

AC AAR79163;  
 XX  
 DT 22-DEC-1995 (first entry)  
 XX  
 DE Partial sequence of bovine alpha 3 chain of type IV collagen.  
 XX  
 KW Type IV collagen; alpha 3 chain.  
 XX  
 OS Bos taurus.  
 XX  
 PN US5424408-A.  
 XX  
 PD 13-JUN-1995.  
 XX  
 PF 30-NOV-1990; 90US-0621091.  
 XX  
 PR 30-NOV-1990; 90US-0621091.  
 XX  
 (UNIV ) UNIV KANSAS MEDICAL CENT.  
 (UYVA ) UNIV YALE.

PI Hudson BG, Morrison KE, Reeders ST;  
 XX  
 DR WPI; 1995-262631/34.  
 DR N-PSDB; AA096290.  
 XX

PT cDNAs encoding human or bovine alpha-3 type 4 collagen peptide(s)  
 PT useful for detection and therapeutic removal of antibodies associated  
 PT with Goodpasture syndrome

Disclosure; Columns 5-8; 33pp; English.

CC Using the PCR with primers derived from each end of the known 27 AA  
 CC residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp  
 CC bovine genomic fragment was amplified. This fragment was then used  
 CC to a bovine lens cDNA library and a 1.5 kb partial cDNA clone was  
 CC obtd. (clone KMC15). This encodes 238 residues of the triple helical  
 CC collagenous domain and all 233 residues of the C-terminal non-  
 CC collagenous (NC1) domain of the alpha 3 (IV) chain. An isolated and  
 CC substantially pure nt. having the sequence in AA096290 is claimed.

Sequence 471 AA;

Query Match 93.7%; Score 986; DB 16; Length 471;

Best Local Similarity 92.6%; Pred. No. 6.2e-103;

Matches 176; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

1 ORAHGODLGTGSCLOPFTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNAPIGR 60  
 DB :|||||  
 DB1 eqahbgd|gltgsc|grftmp|fnc|ndv|cnf|asr|ndys|w|st|pam|pmd|api|gr 340  
 OY 61 ALEPYISRCTYCEGPAIAIAVHSOTTDIPPCPHGWSLWKGFSTIMFTSAGSEGTGALA 120  
 DB :|||||  
 DB1 amep|ysr|ctv|ce|gpa|ia|ia|v|hs|ot|td|ip|pc|ph|g|w|sl|wk|gf|st|im|ft|s|ag|se|gt|ga|la 400  
 OY 121 SPGSLLEPRAPFLERCGRCGRCNYSNSYSFWLASLNPFRFRKPISTYKAGLEKII 180  
 DB :|||||  
 DB1 spgs|cle|e|pr|ap|fl|er|cg|rc|g|rc|n|y|s|n|s|y|f|w|la|sl|np|fr|fr|k|pi|st|y|k|ag|le|k|i|i 460  
 OY 181 SRCQVCMKKR 190  
 DB :|||||  
 DB1 src|qv|cm|k|mr 470

Search completed: March 6, 2002, 06:56:24  
 Job time: 135 sec

OY 181 SRCQVCMKKR 191  
 Db 208 srcqvcmmkrrh 218

## RESULT 7

AAV44171  
 ID AAY44171 standard; Protein: 471 AA.  
 AC AAY44171;

DT 01-FEB-2000 (first entry)

DE Bovine type IV collagen alpha3 chain protein.

KW Recombinant; bovine; alpha3 chain; type IV collagen; detection;  
 KW Goodpasture syndrome; antibody; blood; tissue; human; nephrotoxicity.

OS Bos taurus.

PE US5973120-A.

PD 26-OCT-1999.

PR 07-MAR-1995; 95US-0399889.

PR 30-NOV-1990; 90US-0621091.

PA (UNIV ) UNIV KANSAS MEDICAL CENT.

PI Hudson BG, Reeders ST, Morrison KE;

DR WPI: 1999-610317/52.

DR N-PSDB; AA228774.

PT Isolated alpha 3 chain of type IV collagen polypeptide useful for

PS diagnosis and treatment of Goodpasture syndrome -

XX Claim 1; Column 31-34; 27pp; English.

CC This sequence represents a recombinant bovine alpha3 chain of type IV  
 CC collagen polypeptide. The sequence corresponds to the 238 amino acids  
 CC of the C-terminal end of the triple helical domain and all 233 amino  
 CC acids of the C-terminal non-collagenous domain. Alpha3 chain collagen  
 CC polypeptides are useful for detecting Goodpasture antibodies in blood  
 CC or tissue from a human patient and for treating Goodpasture syndrome,  
 CC especially by neutralising the antibodies in the blood. The polypeptides  
 CC also have a nephrotoxic activity.

XX Sequence 471 AA;

Query Match 93.9%; Score 988; DB 20; Length 471;  
 Best Local Similarity 93.2%; Pred. No. 3.7e-103;

Matches 177; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

OY 1 ORAHGDLGTGSCLORETFMPLFCNVNDVCFNFSRNDYSYWLSTPALMNMNAPITGR 60

Db 281 egahgqdlgtlgsclqrlftmplfncndvcnfasrndysywlstpalmpmdapltgr 340

OY 61 ALEPYISRCTVCEGPALAIIVHSQTDIPCPHGMISIMKGFSTIMTSAGSEGTGALA 120

Db 341 alepyisrcvcegpalaiahvsgtdipcpagwistwkgfstimtsagseagaala 400

OY 121 SPGSCLEEFRAFPLECHGRTGTCNVYSNSYFWLASLNPFRMRKPIPTVKKAGLEKII 180

Db 401 spgscleefraspflechgrygtcnynsyfwlasldpkrmfkpiptvkvagelenii 460

OY 181 SRCQVCMKKR 190

Db 461 srcqvcmmkrr 470

## RESULT 8

AAV56783  
 ID AAY56783 standard; Protein: 471 AA.  
 AC AAY56783;

DT 27-MAR-2000 (first entry)

DE Bovine alpha3 type IV collagen.

KW Goodpasture syndrome; type IV collagen; alpha3 chain; bovine.

OS Bos sp.

PE US607980-A.

PD 28-DEC-1999.

PR 07-OCT-1998; 98US-0167364.

PR 30-NOV-1990; 90US-0621091.

PR 07-MAR-1995; 95US-0399889.

PA (UNIV ) UNIV KANSAS MEDICAL CENT.

PI Hudson BG, Reeders ST, Morrison KE;

DR WPI: 2000-096371/08.

DR N-PSDB; AA246728.

PT Diagnosing and treating Goodpasture syndrome using a peptide derived

PS from type IV collagen -

XX Disclosure; Columns 19-24; 26pp; English.

CC The invention provides a method of detecting Goodpasture antibodies  
 CC in the fluid of a patient by contacting it with a peptide comprising at  
 CC most 218 amino acids of the human alpha3 chain type IV collagen that  
 CC contains the fragment shown in AAY56785. The methods are useful for the  
 CC diagnosis and treatment of Goodpasture syndrome. The present sequence  
 CC represents the bovine alpha3 chain of type IV collagen.

XX Sequence 471 AA;

Query Match 93.9%; Score 988; DB 21; Length 471;  
 Best Local Similarity 93.2%; Pred. No. 3.7e-103;

Matches 177; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

OY 1 ORAHGDLGTGSCLORETFMPLFCNVNDVCFNFSRNDYSYWLSTPALMNMNAPITGR 60

Db 281 egahgqdlgtlgsclqrlftmplfncndvcnfasrndysywlstpalmpmdapltgr 340

OY 61 ALEPYISRCTVCEGPALAIIVHSQTDIPCPHGMISIMKGFSTIMTSAGSEGTGALA 120

Db 341 alepyisrcvcegpalaiahvsgtdipcpagwistwkgfstimtsagseagaala 400

OY 121 SPGSCLEEFRAFPLECHGRTGTCNVYSNSYFWLASLNPFRMRKPIPTVKKAGLEKII 180

Db 401 spgscleefraspflechgrygtcnynsyfwlasldpkrmfkpiptvkvagelenii 460

OY 181 SRCQVCMKKR 190

Db 461 srcqvcmmkrr 470

## RESULT 9

AAV79163  
 ID AAR79163 standard; Protein: 471 AA.

XX



RESULT 4  
AAV31993  
ID AAV31993 standard; Protein; 268 AA.  
XX  
AC AAV31993;  
XX  
DT 05-JAN-2000 (first entry)  
XX  
DE Type IV collagen NCI domain alpha-3 monomer.  
XX  
KW Type IV collagen; NCI domain; non-collagenous domain; human;  
KW angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;  
KW rheumatoid arthritis; retinal neovascularization;  
KW choroidal neovascularization; macular degeneration;  
KW corneal neovascularization; retinopathy of prematurity;  
KW corneal graft rejection; neovascular glaucoma;  
KW retrolental fibroplasia; epidemic keratoconjunctivitis;  
KW vitamin A deficiency; contact lens overwear; atopic keratitis;  
KW superior limbic keratitis; pterygium keratitis sicca; sogrens;  
KW acne rosacea; phlyctenulosis; syphilis; Mycobacteria infection;  
KW lipid degeneration; chemical burn; ulcer; herpes simplex infection;  
KW Herpes zoster infection; protozoan infection; Kaposi's sarcoma;  
KW Mooren ulcer; Terrien's marginal degeneration;  
KW marginal keratolysis; trauma; systemic lupus; polyarteritis;  
KW Wegener's sarcoidosis; scleritis; Steven's Johnson disease;  
KW radial keratotomy; sickle cell anaemia; sarcoid;  
KW pseudoxanthoma elasticum; Paget's disease; vein occlusion;  
KW artery occlusion; carotid obstructive disease; chronic uveitis;  
KW chronic vitritis; Lyme's disease; Eales disease; Bechets disease;  
KW myopia; optic pit; Stargart's disease; pars planitis;  
KW chronic retinal detachment; hyperviscosity syndrome; toxoplasmosis;  
KW post-laser complication; fibrovascular tissue proliferation;  
KW haemangioma; Osler-Weber-Rendu; AIDS; ocular neovascular disease;  
KW osteoarthritis; chronic inflammation; Crohn's disease;  
KW ulcerative colitis; psoriasis; atherosclerosis; pemphigoid.  
XX  
XX Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..17  
FT /note= "BM40 signal peptide"  
FT Protein 18..268  
FT /note= "mature protein"  
FT Peptide 18..25  
FT /note= "affinity tag"  
FT Protein 26..268  
FT /note= "NCI alpha-3 monomer"  
XX  
PN MO949885-A2.  
XX  
PD 07-OCT-1999.  
XX  
XX 26-MAR-1999; 99MO-US06445.  
XX  
XX 27-MAR-1998; 98US-0079783.  
XX  
XX 29-OCT-1998; 98US-0106170.  
XX  
PA (UNITV ) UNIV KANSAS MEDICAL CENT.  
XX  
PI Hudson BG, Sarraas MP;  
XX  
DR WPI: 1999-601297/51.  
DR N-PSDB; AA220091.  
XX  
PT Inhibition of angiogenesis with non-collagenous alpha chain monomer  
PT useful for treating e.g. tumor growth or metastasis,  
XX neovascularisation, etc  
XX  
PS Disclosure; Fig 17c; 56pp; English.  
XX

CC This sequence represents a recombinant type IV collagen  
CC non-collagenous (NCI) domain alpha-3 polypeptide composed of a  
CC BM40 signal sequence (which is cleaved from the mature protein) to  
CC facilitate protein secretion, and a mature protein comprising an  
CC affinity tag (facilitates purification and identification of the  
CC material) and the alpha-1 chain monomer. The invention provides  
CC methods and kits for inhibiting angiogenesis, tumour growth and  
CC metastasis, and endothelial cell interaction with the extracellular  
CC matrix, each method comprising contacting the tumour or animal  
CC tissue with 1 or more isolated type IV collagen NCI alpha chain  
CC monomer(s) selected from the group consisting of alpha-1, alpha-2,  
CC alpha-3 and alpha-6 NCI chain monomers (see AAV31991-96). The  
CC monomers can be produced via recombinant protein expression. The  
CC polynucleotides and polypeptides are used to treat an angiogenesis-  
CC mediated disorder or condition, especially selected from solid and  
CC blood-borne tumours, diabetic retinopathy, rheumatoid arthritis,  
CC retinal neovascularization, choroidal neovascularization, macular  
CC degeneration, corneal neovascularization, retinopathy of prematurity,  
CC corneal graft rejection, neovascular glaucoma, retrolental  
CC fibroplasia, epidemic keratoconjunctivitis, vitamin A deficiency,  
CC contact lens overwear, atopic keratitis, superior limbic keratitis,  
CC pterygium keratitis sicca, sogrens, acne rosacea, phlyctenulosis,  
CC syphilis, Mycobacteria infections, lipid degeneration, chemical  
CC burns, bacterial ulcers, fungal ulcers, herpes simplex infections,  
CC herpes zoster infections, protozoan infections, Kaposi's sarcoma,  
CC Mooren ulcer, Terrien's marginal degeneration, marginal keratolysis,  
CC trauma, systemic lupus, polyarteritis, Wegener's sarcoidosis,  
CC scleritis, Steven's Johnson disease, radial keratotomy, sickle cell  
CC anaemia, sarcoid, pseudoxanthoma elasticum, Paget's disease, vein  
CC occlusion, artery occlusion, carotid obstructive disease, chronic  
CC uveitis, myopia, optic pit, Stargart's disease, pars planitis,  
CC disease, myopia, optic pits, Stargart's disease, Bechets  
CC chronic retinal detachment, hyperviscosity syndromes, toxoplasmosis,  
CC post-laser complications, abnormal proliferation of fibrovascular  
CC tissue, haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular  
CC disease, osteoarthritis, chronic inflammation, Crohn's disease,  
CC ulcerative colitis, psoriasis, atherosclerosis, and pemphigoid (all  
CC claimed).  
CC  
XX  
SQ Sequence 268 AA;  
XX  
Query Match 99.5%; Score 1047; DB 20; Length 268;  
Best Local Similarity 99.5%; Pred. No. 3.7e-110;  
Matches 190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ORAHGDLGTLGSLQRTTTPFIFCVNDVNCNRSNDYSYWLSTPALPMNMAPITGR 60  
DB 78 gragqdgldgtlgsclgrftttmpflfcvndvncnlasrndysywlstpalpmnmapitgr 137  
QY 61 ALEPYISRCIVCEGPATAIAVHSQTDIPPCPHGISLWKGFSFIMPTSAGSEGTGQALA 120  
DB 138 alepyisrcivcegpataiavhsgtdlppcpghgislwkgfsfimtsgssegagqala 197  
QY 121 SPGSCLEEFRAFPFLBEHGRCOTWYNSYSFWLASLNPERMRKPIPSYVKALELEKII 180  
DB 198 spgsclleefraspflebhgrctwnysysfwlaslnpermrkpiipsyvkalelekii 257  
QY 181 SROGVCMKKRH 191  
DB 258 srogvcmkrrh 268  
XX  
RESULT 5  
AAV97555  
ID AAV97555 standard; Protein; 268 AA.  
XX  
AC AAV97555;  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Human alpha3(IV)NCI protein sequence.  
XX

Db 115 alepylsrcvcegpalaiahvhsqtdlpcpgwlsiwkgsfifmftsagsegcgala 174  
 QY 121 SPGSCLEEFRAFPFLCHGRCNYSNSYSFWLASLNPRMRKRPISVYKAGLEKII 180  
 Db 175 spgscleefrafpflchgrgcnysnsysfwlaslnpermrfrkripstvkagelekii 234  
 QY 181 SRCQVCMKRRH 191  
 Db 235 srcqvcmkrrh 245

## RESULT 2

AAV44172

ID AAV44172 standard; Protein; 218 AA.

XX AAV44172;

XX 01-FEB-2000 (first entry)

XX Human type IV collagen alpha3 chain protein.

XX Recombinant; bovine; alpha3 chain; type IV collagen; detection;  
 KW Goodpasture syndrome; antibody; blood; tissue; human; nephrotrophism.

XX Homo sapiens.

XX US5973120-A.

XX 26-OCT-1999.

XX 07-MAR-1995; 9505-0399889.

XX 30-NOV-1990; 90US-0621091.

XX (UYVA ) UNIV YALE.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX Hudson BG, Reeders ST, Morrison KE;

XX WPI; 1999-610317/52.

XX N-PSDB; AA28775.

XX Isolated alpha 3 chain of type IV collagen polypeptide useful for

XX diagnosis and treatment of Goodpasture syndrome -

XX Claim 2; Column 35-36; 27pp; English.

XX This sequence represents a recombinant human alpha3 chain of type IV  
 CC collagen polypeptide. The sequence corresponds to the 218 amino acids  
 CC of the C-terminal non-collagenous domain. Alpha3 chain collagens  
 CC polypeptides are useful for detecting Goodpasture antibodies in blood  
 CC or tissue from a human patient and for treating Goodpasture syndrome,  
 CC especially by neutralising the antibodies in the blood. The polypeptides  
 CC also have a nephrotrophic activity.

XX Sequence 218 AA;

Query Match 99.5%; Score 1047; DB 20; Length 218;  
 Best Local Similarity 99.5%; Pred. No. 2.8e-110;  
 Matches 190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRAHGDDLTGSLCLORTTTPFLFCNVNDVCFNFSNDYSYWLSPALPMNMAPITGR 60  
 Db 28 qrahgddltgslcqlgrfttmplfcnvndvcfnfsndysywlspalpmmapitgr 87  
 QY 61 ALEPYISRCVCEGPALIAIVHSQTDIPCPHGWISLWKGSFIMFTSAGSGTGALA 120  
 Db 88 alepylsrcvcegpalaiahvhsqtdlpcpgwlsiwkgsfifmftsagsegcgala 147  
 QY 121 SPGSCLEEFRAFPFLCHGRCNYSNSYSFWLASLNPRMRKRPISVYKAGLEKII 180

Db 148 spgscleefrafpflchgrgcnysnsysfwlaslnpermrfrkripstvkagelekii 207  
 QY 181 SRCQVCMKRRH 191  
 Db 208 srcqvcmkrrh 218

## RESULT 3

AAV56784

ID AAV56784 standard; Protein; 218 AA.

XX AAV56784;

XX 27-MAR-2000 (first entry)

XX Human alpha3 type IV collagen C-terminal domain.

XX Goodpasture syndrome; type IV collagen; alpha3 chain; human.

XX Homo sapiens.

XX US6007980-A.

XX 28-DEC-1999.

XX 07-OCT-1998; 98US-0167364.

XX 30-NOV-1990; 90US-0621091.

XX 07-MAR-1995; 9505-0399889.

XX (UNIV ) UNIV KANSAS MEDICAL CENT.

XX (UYVA ) UNIV YALE.

XX Hudson BG, Reeders ST, Morrison KE;

XX WPI; 2000-096371/08.

XX N-PSDB; AA246729.

XX Diagnosing and treating Goodpasture syndrome using a peptide derived

XX from type IV collagen -

XX Disclosure; Columns 23-26; 26pp; English.

XX The invention provides a method of detecting Goodpasture antibodies  
 CC in the fluid of a patient by contacting it with a peptide comprising at  
 CC most 218 amino acids of the human alpha3 chain type IV collagen that  
 CC contains the fragment shown in AAV56785. The methods are useful for the  
 CC diagnosis and treatment of Goodpasture syndrome. The present sequence  
 CC represents the carboxy terminal noncollagenous domain of the human  
 CC alpha3 chain of type IV collagen.

XX Sequence 218 AA;

Query Match 99.5%; Score 1047; DB 21; Length 218;  
 Best Local Similarity 99.5%; Pred. No. 2.8e-110;  
 Matches 190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRAHGDDLTGSLCLORTTTPFLFCNVNDVCFNFSNDYSYWLSPALPMNMAPITGR 60  
 Db 28 qrahgddltgslcqlgrfttmplfcnvndvcfnfsndysywlspalpmmapitgr 87  
 QY 61 ALEPYISRCVCEGPALIAIVHSQTDIPCPHGWISLWKGSFIMFTSAGSGTGALA 120  
 Db 88 alepylsrcvcegpalaiahvhsqtdlpcpgwlsiwkgsfifmftsagsegcgala 147  
 QY 121 SPGSCLEEFRAFPFLCHGRCNYSNSYSFWLASLNPRMRKRPISVYKAGLEKII 180  
 Db 148 spgscleefrafpflchgrgcnysnsysfwlaslnpermrfrkripstvkagelekii 207  
 QY 181 SRCQVCMKRRH 191  
 Db 208 srcqvcmkrrh 218

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: March 6, 2002, 06:54:09 ; Search time 54.67 Seconds

(without alignments)  
258.789 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_54\_244

Sequence: 1 ORAHGQDLGTIGSCLQRTT.....KAGELEKTIISRCVCMKKRH 191

## Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Tr number of hits satisfying chosen parameters: 9

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 90%

Listing first 1000 summaries

## Database :

A.Geneseq\_1101.\*  
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1052	100.0	245	21	AAV67942 Human type IV coll
2	1047	99.5	218	20	AAV44172 Human type IV coll
3	1047	99.5	218	21	AAV56784 Human alpha3 type
4	1047	99.5	268	20	AAV31993 Type IV collagen N
5	1047	99.5	268	21	AAV97555 Human alpha3(IV)NC
6	1025	97.4	218	16	AAV79164 Partial sequence o
7	988	93.9	471	20	Bovine type IV col
8	988	93.9	471	21	Bovine alpha3 type
9	986	93.7	471	16	AAV79163 Partial sequence o

## ALIGNMENTS

## RESULT 1

ID AAV67942 standard: Protein; 245 AA.

AC AAV67942:

DT 03-APR-2000 (first entry)

XX Human type IV collagen alpha 3 chain protein sequence SEQ ID NO:10.

XX Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;

XX benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;

XX ocular angiogenesis disease; Osler-Weber Syndrome; telangiectasia;

XX myocardial angiogenesis; plaque neovascularisation; angiodioma;

XX atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;

XX contraception; obesity.

XX Homo sapiens.

XX WO965940-A1.

XX 23-DEC-1999.

XX 17-JUN-1999; 99WO-US13737.

XX 17-JUN-1998; 98US-0089689.

XX 25-MAR-1999; 99US-0126175.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI: 2000-097708/08.

XX N-PSDB: AA57158.

XX Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or

XX 3 chain of type IV collagen used in, e.g. treatment of benign tumors

XX and rheumatoid arthritis -

XX Claim 32: Fig 16B; 117pp; English.

XX The present sequence represents the human type IV collagen alpha 3 chain.

XX The present invention describes an isolated protein chosen from the NC1

XX domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or

XX a fragment, analogue, derivative or mutant, which has anti-angiogenic

XX properties. The anti-angiogenic proteins, multimers and chimeras are

XX useful for inhibiting angiogenic activity in mammalian tissue,

XX especially for treating diseases chosen from angiogenesis-dependent

XX cancers, benign tumours, rheumatoid arthritis, diabetic retinopathy,

XX psoriasis, ocular angiogenesis diseases, Osler-Weber Syndrome,

XX myocardial angiogenesis, plaque neovascularisation, telangiectasia,

XX haemophilic joints, angiodioma, wound granulation, intestinal

XX adhesions, atherosclerosis, scleroderma, hypertrophic scars, cat scratch

XX disease, Helicobacter pylori ulcers, dialysis graft vascular access

XX stenosis, contraception and obesity. The compositions can be used to

XX inhibit a disease characterised by angiogenic activity, in conjunction

XX with radiation therapy, chemotherapy or immunotherapy.

XX Sequence 245 AA:

Query Match 100.0%; Score 1052; DB 21; Length 245;  
Best Local Similarity 100.0%; Pred. No. 8.e-111;

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORAHGQDLGTIGSCLQRTTTPFLFCNVNDYCNFASRNDYSYWLSTPALMPMNAPIITGR 60

Db 55 qrahgqdlgtlgsclqgrtftmpflfcnvndyctnasrndysywlstpalmppmnaplitr 114

OY 61 ALEPYISRCTVCEGPAITAIVHSQTTDIPCPHGWMKMGFSIMFTSAGSECTGALA 120

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 06:54:54 ; Search time 50.56 Seconds  
(without alignments)  
552.571 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_54\_244

Perfect score: 1052

Sequence: 1 QRAHGQDLGTLGSCLORETT.....KAGELEKIISRCQVCMKKRH 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues.

Tc number of hits satisfying chosen parameters: 8

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 90%

Maximum Match 100%

Listing first 1000 summaries

Database :

SPTREMBL\_17.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1052	100.0	245	4 Q9NYC4	Q9nyc4 homo sapien
2	1052	100.0	1670	4 Q9BOT2	Q9bot2 homo sapien
3	1043	99.1	212	6 Q28512	Q28512 macaca mula
4	1000	95.1	212	6 Q28567	Q28567 ovis aries
5	996	94.7	246	11 Q61435	Q61435 mus musculus
6	996	94.7	1669	11 Q9QZS0	Q9qzs0 mus musculus
7	990	94.1	230	11 Q63122	Q63122 rattus norv
8	963	91.5	203	6 Q28682	Q28682 oryctolagus

#### ALIGNMENTS

RESULT 1  
Q9NYC4  
ID Q9NYC4 PRELIMINARY; PRT; 245 AA.  
AC Q9NYC4;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DT TUNSTATIN (FRAGMENT).  
DE Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,  
RA Erickson M.D., Hopper H., Xiao Y., Stillman I.E., Kalluri R.;  
RT "Distinct anti-tumor properties of a type IV collagen domain derived  
from basement membrane.";  
RL J. Biol. Chem. 0:0-0(2000).  
DR EMBL: AF258351; AAF72632.1;  
DR InterPro: IPR001442; C4;  
DR InterPro: IPR000504; RRM.  
DR Pfam: PF01413; C4; 2.  
DR ProDom: PD003923; C4; 2.  
DR SMART: SM00111; C4; 2.  
DR PROSITE: PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
FT NON\_TER 1  
SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 100.0%; Score 1052; DB 4; Length 245;

Best Local Similarity 100.0%; Pred. No. 1.4e-101;

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRAHGQDLGTLGSCLORETTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPITGR 60  
|||||  
DB 55 QRAHGQDLGTLGSCLORETTMPFLFCNVNDVCFASRNDYSYWLSTPALMPMNPITGR 114  
|||||  
QY 61 ALEPYISRCTVCEGPAITAIVHSQTTDIPCPHGWSLWKGFSFIMFTSAGSGTGOALA 120  
|||||  
DB 115 ALEPYISRCTVCEGPAITAIVHSQTTDIPCPHGWSLWKGFSFIMFTSAGSGTGOALA 174  
|||||  
QY 121 SPGSCLEEFPRASPFLECHGRGTCNYNSYFWLASINPERMERKPIPTSPVKAGELEKII 180  
|||||  
DB 175 SPGSCLEEFPRASPFLECHGRGTCNYNSYFWLASINPERMERKPIPTSPVKAGELEKII 234  
|||||  
QY 181 SRCQVCMKKRH 191  
|||||  
DB 235 SRCQVCMKKRH 245

RESULT 2

Q9BOT2

ID Q9BOT2 PRELIMINARY; PRT; 1670 AA.

AC Q9BOT2;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE ALPHA3 TYPE IV COLLAGEN.

GN COL4A3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21064696; PubMed=11134255;

RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,

RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;

RT "Structure of the human type IV collagen gene COL4A3 and mutations in

autosomal Alport syndrome.";

RL J. Am. Soc. Nephrol. 12:97-106(2001).

DR EMBL: AJ288487; CAC36101.1;

DR EMBL: AJ288488; CAC36101.1; JOINED.

DR EMBL: AJ288489; CAC36101.1; JOINED.

DR EMBL: AJ288490; CAC36101.1; JOINED.

DR EMBL: AJ288491; CAC36101.1; JOINED.

DR EMBL: AJ288492; CAC36101.1; JOINED.

DR EMBL: AJ288493; CAC36101.1; JOINED.

DR EMBL; AJ288494; CAC36101.1; JOINED.  
 DR EMBL; AJ288495; CAC36101.1; JOINED.  
 DR EMBL; AJ288496; CAC36101.1; JOINED.  
 DR EMBL; AJ288497; CAC36101.1; JOINED.  
 DR EMBL; AJ288498; CAC36101.1; JOINED.  
 DR EMBL; AJ288499; CAC36101.1; JOINED.  
 DR EMBL; AJ288500; CAC36101.1; JOINED.  
 DR EMBL; AJ288501; CAC36101.1; JOINED.  
 DR EMBL; AJ288502; CAC36101.1; JOINED.  
 DR EMBL; AJ288503; CAC36101.1; JOINED.  
 DR EMBL; AJ288504; CAC36101.1; JOINED.  
 DR EMBL; AJ288505; CAC36101.1; JOINED.  
 DR EMBL; AJ288506; CAC36101.1; JOINED.  
 DR EMBL; AJ288507; CAC36101.1; JOINED.  
 DR EMBL; AJ288508; CAC36101.1; JOINED.  
 DR EMBL; AJ288509; CAC36101.1; JOINED.  
 DR EMBL; AJ288510; CAC36101.1; JOINED.  
 DR EMBL; AJ288511; CAC36101.1; JOINED.  
 DR EMBL; AJ288512; CAC36101.1; JOINED.  
 DR EMBL; AJ288513; CAC36101.1; JOINED.  
 DR EMBL; AJ288514; CAC36101.1; JOINED.  
 DR EMBL; AJ288515; CAC36101.1; JOINED.  
 DR EMBL; AJ288516; CAC36101.1; JOINED.  
 DR EMBL; AJ288517; CAC36101.1; JOINED.  
 DR EMBL; AJ288518; CAC36101.1; JOINED.  
 DR EMBL; AJ288519; CAC36101.1; JOINED.  
 DR EMBL; AJ288520; CAC36101.1; JOINED.  
 DR EMBL; AJ288521; CAC36101.1; JOINED.  
 DR EMBL; AJ288522; CAC36101.1; JOINED.  
 DR EMBL; AJ288523; CAC36101.1; JOINED.  
 DR EMBL; AJ288524; CAC36101.1; JOINED.  
 DR EMBL; AJ288525; CAC36101.1; JOINED.  
 DR EMBL; AJ288526; CAC36101.1; JOINED.  
 DR EMBL; AJ288527; CAC36101.1; JOINED.  
 DR EMBL; AJ288528; CAC36101.1; JOINED.  
 DR EMBL; AJ288529; CAC36101.1; JOINED.  
 DR EMBL; AJ288530; CAC36101.1; JOINED.  
 DR EMBL; AJ288531; CAC36101.1; JOINED.  
 DR EMBL; AJ288532; CAC36101.1; JOINED.  
 DR EMBL; AJ288533; CAC36101.1; JOINED.  
 DR EMBL; AJ288534; CAC36101.1; JOINED.  
 DR EMBL; AJ288535; CAC36101.1; JOINED.  
 DR EMBL; AJ288536; CAC36101.1; JOINED.  
 DR EMBL; AJ288537; CAC36101.1; JOINED.  
 DR EMBL; AJ288538; CAC36101.1; JOINED.  
 KW Collagen.  
 SQ SEQUENCE 1670 AA; 161899 MW; FA7BE4914CA0A6F6 CRC64;

Query Match 100.0%; Score 1052; DB 4; Length 1670;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-100;  
 Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QRAHGDLGTLGSLQRTTMTPLFCNVNDVCFNFSRNDYSYWLSTPALMPNMPIGTR 60  
 DB 1480 QRAHGDLGTLGSLQRTTMTPLFCNVNDVCFNFSRNDYSYWLSTPALMPNMPIGTR 1539  
 QY 61 ALEPYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLAKGFSFMFTSAGSEGTGQALA 120  
 DB 1540 ALEPYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLAKGFSFMFTSAGSEGTGQALA 1599  
 QY 121 SPGSCLEEFRAFPFLCHGRCGTCNYNSYNSYFWLASLNPFMRFRKPIPTVVRAGELEKII 180  
 DB 1600 SPGSCLEEFRAFPFLCHGRCGTCNYNSYNSYFWLASLNPFMRFRKPIPTVVRAGELEKII 1659  
 QY 181 SRCQVCMKKRH 191  
 DB 1660 SRCQVCMKKRH 1670  
 RESULT 3  
 Q28512  
 ID Q2 PRT; 212 AA.

AC Q28512;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).  
 GN COL4A3.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopitheidae; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=RENAL CORTEX;  
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbama I.,  
 RA Mason P.J., Pusey C.D.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; I47280; AAA91861.1; -;  
 DR InterPro; IPR000504; RRM.  
 DR InterPro; IPR001442; C4.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 DR SMART; SM00111; C4; 2.  
 FT NON\_TER 1  
 FT NON\_TER 212  
 SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357E64 CRC64;  
 Query Match 99.1%; Score 1043; DB 6; Length 212;  
 Best Local Similarity 99.0%; Pred. No. 1.1e-100;  
 Matches 189; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 QRAHGDLGTLGSLQRTTMTPLFCNVNDVCFNFSRNDYSYWLSTPALMPNMPIGTR 60  
 DB 22 ERAHGDLGTLGSLQRTTMTPLFCNVNDVCFNFSRNDYSYWLSTPALMPNMPIGTR 81  
 QY 61 ALEPYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLAKGFSFMFTSAGSEGTGQALA 120  
 DB 82 ALEPYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLAKGFSFMFTSAGSEGTGQALA 141  
 QY 121 SPGSCLEEFRAFPFLCHGRCGTCNYNSYNSYFWLASLNPFMRFRKPIPTVVRAGELEKII 180  
 DB 142 SPGSCLEEFRAFPFLCHGRCGTCNYNSYNSYFWLASLNPFMRFRKPIPTVVRAGELEKII 201  
 QY 181 SRCQVCMKKRH 191  
 DB 202 SRCQVCMKKRH 212  
 RESULT 4  
 Q28567  
 ID Q28567; PRELIMINARY; PRT; 212 AA.  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).  
 GN COL4A3.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=RENAL CORTEX;  
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbama I.,  
 RA Mason P.J., Pusey C.D.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; I47282; AAA91904.1; -;  
 DR InterPro; IPR000504; RRM.  
 DR InterPro; IPR001442; C4.



DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
DR SMART; SM00111; C4; 2.  
FT NON\_TER 1 212  
FT NON\_TER 212 212  
SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 95.1%; Score 1000; DB 6; Length 212;  
Best Local Similarity 92.6%; Pred. No. 3.2e-96;  
Matches 176; Conservative 13; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORAHQDGLTGLSCLQRTTTPFPCVGVNVDYCNFASRNDYSYWLSTPALMPMNPITGR 60  
DB 22 EQAHQDGLTGLSCLQRTTTPFPCVGVNVDYCNFASRNDYSYWLSTPALMPMNPITGR 81  
QY 61 ALPEYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGFSEFIMFTSAGSEGTGQALA 120  
DB 82 ALPEYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGFSEFIMFTSAGSEGTGQALA 141  
QY 121 SPGSCLEEFPRASPFLECHGRGTCNYNSYFWLASLNPERMFRKPIPTSTVKAGELEKII 180  
DB 142 SPGSCLEEFPRASPFLECHGRGTCNYNSYFWLASLNPERMFRKPIPTSTVKAGELEKII 201  
QY 181 SRCQVCMKRR 190  
DB 202 SRCQVCMKRR 211

RESULT 5  
Q61435  
ID Q61435 PRELIMINARY; PRT; 246 AA.  
AC Q61435;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE PROCOLLAGEN TYPE IV, ALPHA 3 (COLLAGEN IV ALPHA 3 CHAIN) (FRAGMENT).  
GN COL4A3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C.  
RX MEDLINE=95050957; PubMed=7962065;  
RA Miner J.H.; Sanes J.R.;  
R "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal  
laminae: sequence, distribution, association with laminins, and  
developmental switches.";  
RL J. Cell Biol. 127:879-891(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C.  
RA Miner J.H.;  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C.  
RA Miner J.H.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; 235166; CAA84529.1;  
DR MGD; MGI:104688; Col4a3.  
DR InterPro; IPR001442; C4.  
DR InterPro; IPR000504; RRM.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; C4; 2.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
FT NON\_TER 1  
SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;

Query Match 94.7%; Score 996; DB 11; Length 246;  
Best Local Similarity 93.2%; Pred. No. 9.7e-96;  
Matches 178; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 1 ORAHQDGLTGLSCLQRTTTPFPCVGVNVDYCNFASRNDYSYWLSTPALMPMNPITGR 60  
DB 56 KRAHQDGLTGLSCLQRTTTPFPCVGVNVDYCNFASRNDYSYWLSTPALMPMNPISGR 115  
QY 61 ALPEYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGFSEFIMFTSAGSEGTGQALA 120  
DB 116 ALPEYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGFSEFIMFTSAGSEGTGQALA 175  
QY 121 SPGSCLEEFPRASPFLECHGRGTCNYNSYFWLASLNPERMFRKPIPTSTVKAGELEKII 180  
DB 176 SPGSCLEEFPRASPFLECHGRGTCNYNSYFWLASLNPERMFRKPIPTSTVKAGELEKII 235  
QY 181 SRCQVCMKRR 191  
DB 236 SRCQVCMKRR 246

RESULT 6  
Q9QZS0  
ID Q9QZS0 PRELIMINARY; PRT; 1669 AA.  
AC Q9QZS0;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE ALPHA 3 COLLAGEN IV.  
GN COL4A3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-KIDNEY.  
RX MEDLINE=20005934; PubMed=10534397;  
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,  
Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;  
R "Insertional mutation of the collagen genes col4a3 and col4a4 in a  
mouse model of alport syndrome.";  
RL Genomics 61:113-124(1999).  
DR EMBL; AF169387; AAD50449.1;  
DR MGD; MGI:104688; Col4a3.  
DR InterPro; IPR001442; C4.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 21.  
DR ProDom; PD003923; C4; 2.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
KW Collagen.  
SQ SEQUENCE 1669 AA; 161769 MW; 30976E59739A47B2 CRC64;

Query Match 94.7%; Score 996; DB 11; Length 1669;  
Best Local Similarity 93.2%; Pred. No. 7.8e-95;  
Matches 178; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 1 ORAHQDGLTGLSCLQRTTTPFPCVGVNVDYCNFASRNDYSYWLSTPALMPMNPITGR 60  
DB 1479 KRAHQDGLTGLSCLQRTTTPFPCVGVNVDYCNFASRNDYSYWLSTPALMPMNPISGR 1538  
QY 61 ALPEYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGFSEFIMFTSAGSEGTGQALA 120  
DB 1539 ALPEYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGFSEFIMFTSAGSEGTGQALA 1598  
QY 121 SPGSCLEEFPRASPFLECHGRGTCNYNSYFWLASLNPERMFRKPIPTSTVKAGELEKII 180  
DB 1599 SPGSCLEEFPRASPFLECHGRGTCNYNSYFWLASLNPERMFRKPIPTSTVKAGELEKII 1658

QY 181 SRCQVCMKKRH 191  
 |||||  
 Db 1659 SRCQVCMKKRH 1669

## RESULT 7

Q63122 PRELIMINARY; PRT; 230 AA.  
 AC Q63122;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).  
 GN COL4A3.

OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 QV NCBI\_TaxID=10116;  
 [1]

## SEQUENCE FROM N.A.

RC STRAIN-SPRAGUE-DAWLEY; TISSUE-KIDNEY CORTEX;  
 RA Turner N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L47281; AAB72238.2;  
 DR InterPro; IPR000504; RRM.  
 DR InterPro; IPR001442; C4.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 DR SMART; SM00111; C4; 2.  
 DR NON\_TER 1  
 FT NON\_TER 230  
 SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match 94.1%; Score 990; DB 11; Length 230;  
 Best Local Similarity 92.1%; Pred. No. 3.8e-95;  
 Matches 176; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY 1 QRAHGDLGTLGSLQRFTHMPFCNVNVCNFSRNDYSYWLSTPALMPMNPITGR 60  
 : |||||  
 Db 40 EHAHGDLGTLGSLQRFTHMPFCNVNVCNFSRNDYSYWLSTPALMPMNPITGR 99  
 : |||||  
 QY 61 ALEPYISRCTVCGPAIAIAVHSQTTDIPCPHGWSLWKGFSEFIMFTSAGSEGTGOALA 120  
 : |||||  
 100 ALEPYISRCTVCGPAIAIAVHSQTTDIPCPHGWSLWKGFSEFIMFTSAGSEGTGOALA 159  
 : |||||  
 QY 121 SPGSCLEEFPRASFLECHGRGTCNYNSYSFWLASLNPERMFRKPIPTSVKAGELEKII 180  
 : |||||  
 Db 160 SPGSCLEEFPRASFLECHGRGTCNYNSYSFWLASLNPERMFRKPIPTSVKAGDLEKII 219  
 : |||||  
 QY 181 SRCQVCMKKRH 191  
 |||||  
 Db 220 SRCQVCMKKRH 230

## RESULT 8

Q28682 PRELIMINARY; PRT; 203 AA.  
 AC Q28682;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).  
 GN COL4A3.

OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQ

RC TISSUE=RENAL CORTEX;  
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,  
 RA Mason P.J., Pusey C.D.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L47283; AAA91893.1;  
 DR InterPro; IPR000504; RRM.  
 DR InterPro; IPR001442; C4.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD003923; C4; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 DR SMART; SM00111; C4; 2.  
 FT NON\_TER 1  
 FT NON\_TER 203  
 SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 91.5%; Score 963; DB 6; Length 203;  
 Best Local Similarity 95.1%; Pred. No. 2.1e-92;  
 Matches 173; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 1 QRAHGDLGTLGSLQRFTHMPFCNVNVCNFSRNDYSYWLSTPALMPMNPITGR 60  
 : |||||  
 Db 22 EHAHGDLGTLGSLQRFTHMPFCNVNVCNFSRNDYSYWLSTPALMPMNPITGR 81  
 : |||||  
 QY 61 ALEPYISRCTVCGPAIAIAVHSQTTDIPCPHGWSLWKGFSEFIMFTSAGSEGTGOALA 120  
 : |||||  
 Db 82 ALEPYISRCTVCGPAIAIAVHSQTTDIPCPHGWSLWKGFSEFIMFTSAGSEGTGOALA 141  
 : |||||  
 QY 121 SPGSCLEEFPRASFLECHGRGTCNYNSYSFWLASLNPERMFRKPIPTSVKAGELEKII 180  
 : |||||  
 Db 142 SPGSCLEEFPRASFLECHGRGTCNYNSYSFWLASLNPERMFRKPIPTSVKAGELEKII 201  
 : |||||  
 QY 181 SR 182  
 ||  
 Db 202 SR 203

Search completed: March 6, 2002, 07:04:44  
 Job time: 590 sec

Wed Mar 6 07:07:44 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 06:54:49 ; Search time 31.16 Seconds  
(without alignments)  
322.691 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_1\_132  
Perfect score: 721  
Sequence: 1 GLKGRDSSSPATWTRGF.....RALEPYISRCTVCGPALAI 132

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

To: number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%  
Maximum Match 100%  
Listing first 1000 summaries

Database : PIR\_68:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	721	100.0	1670	1 CGH03B	collagen alpha 3(I
2	713	98.9	220	2 BA9736	collagen alpha 3(I

## ALIGNMENTS

RESULT 1  
CGH03B  
collagen alpha 3(IV) chain precursor, long splice form - human  
N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form  
C:Species: Homo sapiens (man)  
C:Date: 28-Oct-1994 #sequence\_revision 03-Oct-1995 #text\_change 22-Jun-1999  
C:Accession: A54763; A43928; A44043; A45971; A39786  
R:Matuyama, M.; Lehtonen, A.; Mochizuki, T.; Tryggvason, K.; Reenders, S.T.  
J. Biol. Chem. 269, 23013-23017, 1994  
A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression  
A:Reference number: A54763; M01D:94364994  
A:Accession: A54763  
A:Molecule type: mRNA  
A:Residues: 1-1670 <MAR>  
A:Cross-references: GB:X80031; NID:9577563; PID:9577564  
A:Experimental source: Kidney  
R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
J. Clin. Invest. 89, 592-601, 1992  
A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the at  
A:Reference number: A43928; M01D:92147878  
A:Accession: A43928  
A:Molecule type: mRNA

A:Residues: 1331-1524, 'I', 1526-1670 <TUR>  
A:Cross-references: GB:M81379  
A:Experimental source: Kidney  
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 267, 19780-19784, 1992  
A:Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpas  
tion.  
A:Reference number: A44043; M01D:93015826  
A:Accession: A44043  
A:Molecule type: DNA; mRNA  
A:Residues: 1386-1670 <QUI>  
A:Cross-references: GB:M92993; NID:9177895; PIDN:AAA21610.1; PID:9177896  
A:Note: sequence extracted from NCBI backbone (NCBIP:115597)  
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 269, 17358, 1994  
A:Reference number: A44738; M01D:94274734  
A:Contents: annotation; extratum: correction to intronic sequence in A44043  
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 268, 12090-12094, 1993  
A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
A:Reference number: A45971; M01D:93280184  
A:Accession: A45971  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1427-1444 <BER>  
A:Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly id  
R:Morrison, K.E.; Matuyama, M.; Yang-Feng, T.L.; Reenders, S.T.  
Am. J. Hum. Genet. 49, 545-554, 1991  
A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain  
A:Reference number: A39786; M01D:91353570  
A:Accession: A39786  
A:Molecule type: mRNA  
A:Residues: 1453-1593, 'A', 1595-1670 <MOR>  
A:Cross-references: GB:S55790; NID:9234418; PIDN:AAB19637.1; PID:9234419  
A:Comment: Prolines and lysines at the third position of the tripeptide repeating uni  
ed and subsequently O-glycosylated.  
C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epit  
C:Genetics:  
A:Gene: GDB:COL4A3  
A:Cross-references: GDB:128351; OMIM:120070  
A:Map position: 2q36-2q37  
A:Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete  
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands w  
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha  
monomer trimers amino-terminal domains (with disulfide and desmosine cross-links), dimeri  
er associations in the interrupted helical domain (with disulfide and desmosine cross  
C:Function:  
A:Description: minor structural component of extracellular basement membrane in kidne  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extra  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MA  
F:29-42/Domain: amino-terminal nonhelical, NHI <NHI>  
F:43-1438/Region: interrupted helical  
F:91-793/Region: cell attachment (R-G-D) motif  
F:996-998/Region: cell attachment (R-G-D) motif  
F:1154-1156/Region: cell attachment (R-G-D) motif  
F:1306-1308/Region: cell attachment (R-G-D) motif  
F:1345-1347/Region: cell attachment (R-G-D) motif  
F:1432-1434/Region: cell attachment (R-G-D) motif  
F:1439-1670/Domain: carboxyl-terminal nonhelical, NCI <NCI>  
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>  
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>  
F:31.33, 39, 41, 125, 422, 476, 479, 682, 722, 809, 1387/Disulfide bonds: Interchain #status pr  
F:253/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:1460-1548, 1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
F:1505-1511, 1616-1622/Disulfide bonds: #status predicted  
F:1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 100.0%; Score 721; DB 1; Length 1670;  
Best Local Similarity 100.0%; Prod. No. 1.8e-65;  
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: March 6, 2002, 06:54:49  
 Job time: 40 sec

QY 1 GLKGRKDDSGSPATWTRGTFVTRHSGQTALPSCPECTVLYSGSFLFYQGNORAHGOD 60  
 |||||||  
 Db 1427 GLKGRKDDSGSPATWTRGTFVTRHSGQTALPSCPECTVLYSGSFLFYQGNORAHGOD 1486  
 QY 61 LGTIGSCLQRTTMRPFLEFCNVNDVCNFRASNDYSYWLSTPALMPMNAPIITGRALEPYIS 120  
 |||||||  
 Db 1487 LGTIGSCLQRTTMRPFLEFCNVNDVCNFRASNDYSYWLSTPALMPMNAPIITGRALEPYIS 1546  
 QY 121 RCTVCEGPALAI 132  
 |||||||  
 Db 1547 RCTVCEGPALAI 1558

## RESULT 2

B49736  
 collagen alpha 3(IV) chain, medium splice form - human (fragment)  
 N:Contains: collagen alpha 3(IV) chain, splice form GP-V  
 Species: Homo sapiens (man)  
 Date: 03-May-1994 #sequence\_revision 12-Nov-1999 #text\_change 17-Mar-2000  
 Accession: B49736; S69111  
 R: Feng, L.; Xia, Y.; Wilson, C.B.  
 J. Biol. Chem. 269, 2342-2348, 1994  
 A: Title: Alternative splicing of the NC1 domain of the human alpha3(IV) collagen gene.  
 A: Reference number: A49736; MUID: 94124597  
 A: Accession: B49736  
 A: Status: nucleic acid sequence not shown  
 A: Molecule type: mRNA  
 A: Residues: 169-220 <FEN1>  
 A: Accession: D49736  
 A: Status: nucleic acid sequence not shown; translation not shown  
 A: Molecule type: mRNA  
 A: Residues: 22-220 <FEN2>  
 A: Cross-references: GB:002519; NID:9409106; PIDN:AAA18942.1; PID:9409107  
 A: Note: this is the conceptual translation of the nucleic acid submitted to GenBank  
 R: Penades, J. R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V. J.; Cervera, J.; Wild  
 Eur. J. Biochem. 229, 754-760, 1995  
 A: Title: Characterization and expression of multiple alternatively spliced transcripts c  
 utonitigen and one of its alternative forms.  
 A: Reference number: S69111; MUID: 95278230  
 A: Molecule type: mRNA  
 A: Accession: S69111  
 A: Residues: 1-45, 169-204, 'L', 206-220 <PEN>  
 A: Comment: For the complete sequence of the long splice form, see PIR:CGH03B.  
 C: Genetics:  
 A: Gene: GDB:COL4A3  
 A: Cross-references: GDB:128351; OMIM:120070  
 P position: 2936-2937  
 P family: collagen alpha 1(IV) chain  
 C: Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel  
 l; 1-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status predi  
 F; 1-45, 169-220/Product: collagen alpha 3(IV) chain, splice form GP-V (fragment) #status  
 F; 22-220/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
 F; 34-134/Domain: collagen IV carboxyl-terminal repeat <CT1>

Query Match 98.9%; Score 713; DB 2; Length 220;

Best Local Similarity 99.2%; Pred. No. 1.4e-65;

Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLKGRKDDSGSPATWTRGTFVTRHSGQTALPSCPECTVLYSGSFLFYQGNORAHGOD 60  
 |||||||  
 Db 10 GLKGRKDDSGSPATWTRGTFVTRHSGQTALPSCPECTVLYSGSFLFYQGNORAHGOD 69  
 QY 61 LGTIGSCLQRTTMRPFLEFCNVNDVCNFRASNDYSYWLSTPALMPMNAPIITGRALEPYIS 120  
 |||||||  
 Db 70 LGTIGSCLQRTTMRPFLEFCNVNDVCNFRASNDYSYWLSTPALMPMNAPIITGRALEPYIS 129  
 QY 121 RCTVCEGPALAI 132  
 |||||||  
 Db 130 RCTVCEGPALAI 141

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 07:05:08 ; Search time 18.75 Seconds  
(without alignments)  
258.120 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_1\_132

Perfect score: 721  
Sequence: 1 GLKGRGDSGSPATWTTRGF.....RALEPYISRCTVCEGPAIAI 132

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 100059 segs, 36664827 residues

T number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%  
Maximum Match 100%  
Listing first 1000 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	721	100.0	1670	1 CA34_HUMAN	Q01955 homo sapien

## ALIGNMENTS

RESULT 1  
CA34\_HUMAN STANDARD: PRT; 1670 AA.  
Q01955:  
D 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE COLLAGEN ALPHA 3(IV) CHAIN PRECURSOR (GOODPASTURE ANTIGEN).  
GN COL4A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=94364994; PubMed=8083201;  
RA Maruyama M., Leinonen A., Mochizuki T., Trygvasen K., Reiders S.T.;  
RT "Complete primary structure of the human alpha 3(IV) collagen chain in  
human tissues."  
RL J. Biol. Chem. 269:23013-23017(1994).  
RN [2]  
RP REVISIONS.  
RA Leinonen A.;  
RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=93015826; PubMed=1400291;  
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;  
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the  
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
RT antigenic region at the triple helix/NC1 domain junction."  
RL J. Biol. Chem. 267:19780-19784(1992).  
RN [4]  
RP SEQUENCE OF 1453-1670 FROM N.A.  
RX MEDLINE=91353570; PubMed=1882840;  
RA Morrison K.E., Maruyama M., Yang-Feng T.L., Reiders S.T.;  
RT "Sequence and localization of a partial cDNA encoding the human alpha  
RT 3 chain of type IV collagen."  
RL Am. J. Hum. Genet. 49:545-554(1991).  
RN [5]  
RP SEQUENCE OF 1331-1670 FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=92147878; PubMed=1737849;  
RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
RA Pusey C.D.;  
RT "Molecular cloning of the human Goodpasture antigen demonstrates it  
RT to be the alpha 3 chain of type IV collagen."  
RL J. Clin. Invest. 89:592-601(1992).  
RN [6]  
RP SEQUENCE OF 1644-1670 FROM N.A.  
RC TISSUE=Kidney;  
RA Ding J.;  
RN Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.  
RC TISSUE=Kidney;  
RX MEDLINE=94124597; PubMed=8294492;  
RA Feng L., Xia Y., Wilson C.B.;  
RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)  
RT collagen gene. Differential expression of mRNA transcripts that  
RT predict three protein variants with distinct carboxyl regions."  
RL J. Biol. Chem. 269:2342-2348(1994).  
RN [8]  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=98196854; PubMed=9537506;  
RA Monota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioke H.,  
RA Nimomura Y.;  
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
RT 2q36."  
RL FEBS Lett. 424:11-16(1998).  
RN [9]  
RP ALTERNATIVE SPLICING.  
RX MEDLINE=93280184; PubMed=8505332;  
RA Bernal D., Quinones S., Saus J.;  
RT "The human mRNA encoding the Goodpasture antigen is alternatively  
RT spliced."  
RL J. Biol. Chem. 268:12090-12094(1993).  
RN [10]  
RP VARIANT PRO-1474.  
RX MEDLINE=95078827; PubMed=7987301;  
RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
RA Barrientos A., Monens L.A.H., van Oost B.A., Brunner H.G.,  
RA Reiders S.T., Smeets H.J.M.;  
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
RT recessive Alport syndrome."  
RL Hum. Mol. Genet. 3:1269-1273(1994).  
RN [11]  
RP VARIANTS AS, AND VARIANTS.  
RX MEDLINE=21064696; PubMed=11134255;  
RA Heidt L., Attardel C., Forestier L., Cohen-Solal L., Mollet G.,  
RA Gutierrez B., Stavrou C., Gubler M.C., Antigone C.;  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
RT autosomal Alport syndrome."  
RL J. Am. Soc. Nephrol. 12:97-106(2001).  
CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A "CHICKEN-WIRE"  
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/

CC NIDODEN.  
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS: 1 (SHOWN HERE), 2/V AND  
 CC 3/L5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR  
 CC C-TERMINAL NCI DOMAINS.  
 CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE  
 CC COLLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,  
 CC COCHLEA, LUNG AND BRAIN.  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL  
 CC N-LINKED GLYCOSYLATION SITE.  
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -1- DISEASE: ANTIBODIES AGAINST THE NCI DOMAIN OF ALPHA3(IV) MEDIATE  
 CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS  
 CC CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.  
 CC -1- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I  
 CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY  
 CC GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,  
 CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN  
 CC MALES AND FEMALES.  
 CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X80031; CAA56335.1; -  
 CC EMBL: M92993; AAA21610.1; -  
 CC EMBL: S55790; AAB19637.1; -  
 CC EMBL: M81379; AAA51556.1; -  
 CC EMBL: L08650; AAA52044.1; -  
 CC EMBL: U02519; AAA18942.1; -  
 CC EMBL: U02520; AAA18943.1; -  
 CC EMBL: AB008495; BAA25064.1; -  
 CC MIM: 120070; -  
 CC MIM: 203780; -  
 CC MIM: 233450; -  
 CC InterPro: IPR001442; C4.  
 CC InterPro: IPR000087; Collagen.  
 CC Pfam: PF01413; C4; 2.  
 CC Pfam: PF01391; Collagen; 21.  
 CC ProDom: PD003923; C4; 2.  
 CC SMART: SM00111; C4; 2.  
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 CC Glycoprotein; Basement membrane; Collagen; Signal; Cell adhesion;  
 CC Alternative splicing; Polymorphism; Phosphorylation; Disease mutation;  
 CC Alport syndrome.  
 CC  
 CC FT SIGNAL 1 28  
 CC FT CHAIN 29 1670  
 CC FT DOMAIN 29 42  
 CC FT DOMAIN 43 1438  
 CC FT DOMAIN 1439 1670  
 CC FT DOMAIN 1427 1444  
 CC FT SITE 1426 1427  
 CC FT SITE 791 793  
 CC  
 CC POTENTIAL.  
 CC COLLAGEN ALPHA 3(IV) CHAIN.  
 CC 7S DOMAIN.  
 CC TRIPLE-HELICAL REGION.  
 CC NONHELICAL REGION (NC1) (GOODPASTURE  
 CC ANTIGEN) (BY SIMILARITY).  
 CC EPITOPE (RECOGNIZED BY GOODPASTURE  
 CC ANTIBODIES).  
 CC CLEAVAGE (BY COLLAGENASE)  
 CC (BY SIMILARITY).  
 CC CELL ATTACHMENT SITE (POTENTIAL).

QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB
1	1427	61	1487	121	1547	121	1547	121	1547	121	1547
GLKGRDSSGPAWTRGCVFTRHSOTTAIPSCPEGTVPYSGFSLFVQGNRAHGOD	60	LGTLGSLQRTTTPFLFCNVNDVNCNFSARNDSYWLSTPALMNNAPITGRALPEYIS	120	RCTVCEGPALAI	132	RCTVCEGPALAI	132	RCTVCEGPALAI	132	RCTVCEGPALAI	132
1427	1486	1487	1486	1547	1558	1547	1558	1547	1558	1547	1558

Search completed: March 6, 2002, 07:05:09  
 Job time: 585 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 07:04:44 ; Search time 50.56 Seconds  
(without alignments)  
381.882 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_1\_132

Perfect score: 721  
Sequence: 1 GLGKRGDSGSPATWTRGF.....RALEPYISRCTVCEGPAIAI 132

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

TC number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%  
Maximum Match 100%

Listing first 1000 summaries

Database :

SPTREMBL\_17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	721	100.0	245	09NYC4	09NYC4 homo sapien
2	721	100.0	1670	09BOT2	09bot2 homo sapien

#### ALIGNMENTS

RESULT 1  
09NYC4 PRELIMINARY: PRT: 245 AA.  
AC 09NYC4:  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE TUMSTATIN (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

RN [1]  
RP SEQUENCE FROM N.A.  
RA Maeshima Y., Colorado P.C., Torre A., Hollhaus K.A., Grunkemeyer J.A.,  
RA Erickson M.D., Hopfer H., Xiao Y., Stillman I.E., Kalluri R.,  
RT "Distinct anti-tumor properties of a type IV collagen domain derived  
from basement membrane."  
RL J. Biol. Chem. 0:0-0(2000).  
DR EMBL; AF258351; AAF72632.1; -  
DR InterPro; IPR001442; C4.  
DR InterPro; IPR000504; RRM.  
DR Pfam; PF01413; C4; 2.  
DR ProDom; PD003923; C4; 2.  
DR SMART; SM00111; C4; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
FT NON\_TER  
SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 100.0%; Score 721; DB 4; Length 245;  
Best Local Similarity 100.0%; Prod. No. 2.4e-72;  
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GLGKRGDSGSPATWTRGFVTRHSQTAIPSCPECTVPLVYGSFSLFYQGNORAHGOD 60  
|||||  
DB 2 GLGKRGDSGSPATWTRGFVTRHSQTAIPSCPECTVPLVYGSFSLFYQGNORAHGOD 61  
OY 61 LGTLGSCLOREFTTTPFLFCNVNDVNCNPNASRNDYSYMLSTPALPMNMAPITGALPEYIS 120  
|||||  
DB 62 LGTLGSCLOREFTTTPFLFCNVNDVNCNPNASRNDYSYMLSTPALPMNMAPITGALPEYIS 121  
OY 121 RCTVCEGPAIAI 132  
|||||  
DB 122 RCTVCEGPAIAI 133

RESULT 2  
09BOT2 PRELIMINARY: PRT: 1670 AA.  
AC 09BOT2:  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE ALPHA3 TYPE IV COLLAGEN.  
GN COL4A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21064696; PubMed=11134255;  
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,  
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.,  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
the autosomal Alport syndrome."  
RL J. Am. Soc. Nephrol. 12:97-106(2001).  
DR EMBL; AJ288487; CAC36101.1; -  
DR EMBL; AJ288488; CAC36101.1; JOINED.  
DR EMBL; AJ288489; CAC36101.1; JOINED.  
DR EMBL; AJ288490; CAC36101.1; JOINED.  
DR EMBL; AJ288491; CAC36101.1; JOINED.  
DR EMBL; AJ288492; CAC36101.1; JOINED.  
DR EMBL; AJ288493; CAC36101.1; JOINED.  
DR EMBL; AJ288494; CAC36101.1; JOINED.  
DR EMBL; AJ288495; CAC36101.1; JOINED.  
DR EMBL; AJ288496; CAC36101.1; JOINED.  
DR EMBL; AJ288497; CAC36101.1; JOINED.  
DR EMBL; AJ288498; CAC36101.1; JOINED.  
DR EMBL; AJ288499; CAC36101.1; JOINED.  
DR EMBL; AJ288500; CAC36101.1; JOINED.  
DR EMBL; AJ288501; CAC36101.1; JOINED.  
DR EMBL; AJ288502; CAC36101.1; JOINED.  
DR EMBL; AJ288503; CAC36101.1; JOINED.

DR EMBL; AJ288504; CAC36101.1; JOINED.  
DR EMBL; AJ288505; CAC36101.1; JOINED.  
DR EMBL; AJ288506; CAC36101.1; JOINED.  
DR EMBL; AJ288507; CAC36101.1; JOINED.  
DR EMBL; AJ288508; CAC36101.1; JOINED.  
DR EMBL; AJ288509; CAC36101.1; JOINED.  
DR EMBL; AJ288510; CAC36101.1; JOINED.  
DR EMBL; AJ288511; CAC36101.1; JOINED.  
DR EMBL; AJ288512; CAC36101.1; JOINED.  
DR EMBL; AJ288513; CAC36101.1; JOINED.  
DR EMBL; AJ288514; CAC36101.1; JOINED.  
DR EMBL; AJ288515; CAC36101.1; JOINED.  
DR EMBL; AJ288516; CAC36101.1; JOINED.  
DR EMBL; AJ288517; CAC36101.1; JOINED.  
DR EMBL; AJ288518; CAC36101.1; JOINED.  
DR EMBL; AJ288519; CAC36101.1; JOINED.  
DR EMBL; AJ288520; CAC36101.1; JOINED.  
DR EMBL; AJ288521; CAC36101.1; JOINED.  
DR EMBL; AJ288522; CAC36101.1; JOINED.  
DR EMBL; AJ288523; CAC36101.1; JOINED.  
DR EMBL; AJ288524; CAC36101.1; JOINED.  
DR EMBL; AJ288525; CAC36101.1; JOINED.  
DR EMBL; AJ288526; CAC36101.1; JOINED.  
DR EMBL; AJ288527; CAC36101.1; JOINED.  
DR EMBL; AJ288528; CAC36101.1; JOINED.  
DR EMBL; AJ288529; CAC36101.1; JOINED.  
DR EMBL; AJ288530; CAC36101.1; JOINED.  
DR EMBL; AJ288531; CAC36101.1; JOINED.  
DR EMBL; AJ288532; CAC36101.1; JOINED.  
DR EMBL; AJ288533; CAC36101.1; JOINED.  
DR EMBL; AJ288534; CAC36101.1; JOINED.  
DR EMBL; AJ288535; CAC36101.1; JOINED.  
DR EMBL; AJ288536; CAC36101.1; JOINED.  
DR EMBL; AJ288537; CAC36101.1; JOINED.  
DR EMBL; AJ288538; CAC36101.1; JOINED.  
KW Collagen.  
SQ SEQUENCE 1670 AA: 161899 MW: FA7BE4914CA0A6F6 CRC64;

Query Match 100.0%; Score 721; DB 4; Length 1670;  
Best Local Similarity 100.0%; Pred. No. 1.9e-71;  
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLGKRGDSGSPATWTRGFTVRHSQTVAIPSCPEGTVPVLYSGSFLEVOGNORAHGQD 60  
DB 1427 GLGKRGDSGSPATWTRGFTVRHSQTVAIPSCPEGTVPVLYSGSFLEVOGNORAHGQD 1486  
61 LGTLGSCLOREFTMPFLFCVNDVCNFAASNDYSYWLSTPALMPMNAPIITGRALEPYIS 120  
DB 1487 LGTLGSCLOREFTMPFLFCVNDVCNFAASNDYSYWLSTPALMPMNAPIITGRALEPYIS 1546  
QY 121 RCTVCEGPATAI 132  
DB 1547 RCTVCEGPATAI 1558

Search completed: March 6, 2002, 07:04:44  
Job time: 590 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 06:56:24 ; Search time 54.67 Seconds

(without alignments)  
178,849 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_1\_132

Perfect score: 721  
Sequence: 1 GLKGRGDSGSPATWTRGF.....RALEPYISRCYVCEGPAIAI 132

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 90%  
Maximum Match 100%

Listing first 1000 summaries

Database :

A\_Geneseq\_1101:\*

1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT:\*

2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT:\*

3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT:\*

4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT:\*

5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT:\*

6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT:\*

7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT:\*

8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT:\*

9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT:\*

10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT:\*

11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT:\*

12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT:\*

13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT:\*

14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT:\*

15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT:\*

16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT:\*

17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT:\*

18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT:\*

19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT:\*

20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT:\*

21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT:\*

22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	721	100.0	211	21	AAV95918 Human Goodpasture
2	721	100.0	245	21	AAV67942 Human type IV coll
3	700	97.1	268	20	AAV31993 Type IV collagen N
4	700	97.1	268	21	AAV97555 Human alpha3(IV)NC

#### ALIGNMENTS

RESULT 1

ID	XX	AAV95918 standard; Protein; 211 AA.
AC	XX	AAV95918;
DT	XX	20-NOV-2000 (first entry)
DE	XX	Human Goodpasture antigen Deltav.
XX	XX	
KW	XX	Goodpasture antigen: GPDeltav; goodpasture antigen binding protein;
KM	XX	GBP; human; autoimmune disease; apoptosis; cancer; tumour;
KW	XX	therapy.
OS	XX	Homo sapiens.
PN	XX	WO200050607-A2.
PD	XX	31-AUG-2000.
PF	XX	24-FEB-2000; 2000WO-IB00324.
PR	XX	24-FEB-1999; 99US-0121483.
PA	XX	(SAUS/) SAUS J.
PI	XX	Saus J;
XX	XX	
DR	XX	WPI; 2000-572094/53.
DR	XX	N-PSDB; AAA50367.
PT	XX	Novel Goodpasture antigen binding proteins useful for diagnosing and
PT	XX	treating autoimmune disorders, tumor, and preventing cell apoptosis
XX	XX	
XX	XX	Claim 36; Page 151-152; 158pp; English.
XX	XX	
CC	XX	The present sequence is that of human recombinant Goodpasture
CC	XX	antigen (GP) Deltav, i.e. an alternative form of human GP resulting
CC	XX	from splicing out of exon V. The recombinant protein, lacking the
CC	XX	Met-1 residue, was expressed in bacterial pellets using modified
CC	XX	vector pRT15b carrying GPDeltav cDNA (see AAA50367). The invention
CC	XX	relates to novel Goodpasture antigen binding proteins (GPBs, see
CC	XX	AAV95900-11), which bind to and phosphorylate the unique N-terminal
CC	XX	region of human GP, and which are highly expressed in several
CC	XX	autoimmune conditions. Claimed methods for treating an autoimmune
CC	XX	disease, cell apoptosis or a tumour involve modifying the
CC	XX	expression or activity of GPBP, especially using a GP-derived
CC	XX	peptide, such as GPDeltav.
SQ	XX	Sequence 211 AA:
Query Match	100.0%;	Score 721; DB 21; Length 211;
Best Local Similarity	100.0%;	Pred. No. 1.6e-72;
Matches 132; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	GLKGRGDSGSPATWTRGFVTRHSQTALPSCDECTVPLYSFSLFVQGNRAHGQD 60
DB	1	glkkgkrgdsqspatwttrgfvtrhsqtaipscpegtyplysgflfvgngraqgd 60
QY	61	LGITGSLQRFRTMPFLFCNVNDVCNFRASNDYSYTWLSTALMMNNAPTGRALREYIS 120
DB	61	lgtlgsclqrlftmptflfcnvndvcnftrasytwlstpalmnmnapltgralepyis 120
QY	121	RCYVCEGPAIAI 132
DB	121	rclyvcegpaiiai 132

RESULT 2

ID	AAV67942
AC	AAV67942 standard; Protein; 245 AA.
XX	
XX	
XX	



CC facilitate protein secretion, and a mature protein comprising an  
 CC affinity tag (facilitates purification and identification of the  
 CC material) and the alpha-1 chain monomer. The invention provides  
 CC methods and kits for inhibiting angiogenesis, tumour growth and  
 CC metastasis, and endothelial cell interaction with the extracellular  
 CC matrix, each method comprising contacting the tumour or animal  
 CC tissue with 1 or more isolated type IV collagen NCI alpha chain  
 CC monomer(s) selected from the group consisting of alpha-1, alpha-2,  
 CC alpha-3 and alpha-6 NCI chain monomers (see AAY31991-96). The  
 CC monomers can be produced via recombinant protein expression. The  
 CC polynucleotides and polypeptides are used to treat an angiogenesis-  
 CC mediated disorder or condition, especially selected from solid and  
 CC blood-borne tumours, diabetic retinopathy, rheumatoid arthritis,  
 CC retinal neovascularization, choroidal neovascularization, macular  
 CC degeneration, corneal neovascularization, retinopathy of prematurity,  
 CC corneal graft rejection, neovascular glaucoma, retrolental  
 CC fibroplasia, epidemic keratoconjunctivitis, vitamin A deficiency,  
 CC contact lens overwear, atopic keratitis, superior limbic keratitis,  
 CC pterygium keratitis sicca, seborrheic, acne rosacea, phlyctenulosis,  
 CC syphilis, mycobacteria infections, lipid degeneration, chemical  
 CC burns, bacterial ulcers, fungal ulcers, herpes simplex infections,  
 CC herpes zoster infections, protozoan infections, Kaposi's sarcoma,  
 CC Mooren ulcer, Terrien's marginal degeneration, marginal keratolysis,  
 CC trauma, systemic lupus, polyarteritis, Wegener's sarcoidosis,  
 CC scleritis, Steven's Johnson disease, radial keratotomy, sickle cell  
 CC anaemia, sarcoid, pseudoxanthoma elasticum, Paget's disease, vein  
 CC occlusion, artery occlusion, carotid obstructive disease, chronic  
 CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Bechets  
 CC disease, myopia, optic pits, Stargardt disease, pars planitis,  
 CC chronic retinal detachment, hyperviscosity syndromes, toxoplasmosis,  
 CC post-laser complications, abnormal proliferation of fibrovascular  
 CC tissue, haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular  
 CC disease, osteoarthritis, chronic inflammation, Crohn's disease,  
 CC ulcerative colitis, psoriasis, atherosclerosis, and pemphigoid (all  
 CC claimed).

Sequence 268 AA:

Query Match 97.1%; Score 700; DB 20; Length 268;

Best Local Similarity 100.0%; Pred. No. 4.8e-70;

Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KRQDGSPPATWTRGRFVETRHSSQTTAIPSCPECTVPLYSGFLEVOGNRAHQDGLT 64

DB 29 krgdgsppatwtrgrfvtthsgtlaipscpegtvplysgfslfvgngraqbgdgtl 88

QY 65 GSCLORETTMPFLFCNNVNDVCFASRNDYSYWLSTPALMNMNAPITGRALPEYISRCTV 124

D 89 gsclorettmplfrcnnvndvcnfasrndysywlstpalmmnmapitgralepyisrcv 148

QY 125 CEGPAIAI 132

DB 149 cegpaiaai 156

RESULT 4

AAY97555 standard; Protein; 268 AA.

AA97555:

12-FEB-2001 (first entry)

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

Human alpha3(IV)NC1 protein sequence.

KW chronic inflammation; psoriasis; therapy; alpha3(IV)NC1.

XX Homo sapiens.

XX MO200059532-A1.

PN 12-OCT-2000.

PD 31-MAR-2000; 2000WO-US08678.

PF 01-APR-1999; 99US-0127391.

XX (BIOS-) BIOSTRATUM INC.

PA Brooks P, Hudson B;

PI MPI: 2000-664962/64.

DR N-PSDB; AAA90993.

XX Use of antagonists of specific integrin receptors for inhibiting

PT angiogenesis, tumour growth or metastases, or endothelial cell

PT interactions with the extracellular matrix

XX Disclosure: Fig 17c; 78pp; English.

PS This sequence is a human type IV collagen alpha chain monomer,

XX designated alpha3(IV)NC1. The invention relates to a method for

XX inhibiting angiogenesis, tumour growth or metastases, or endothelial cell

XX interactions with the extracellular matrix, comprising contacting the

XX cells or tissue with a polypeptide composition containing antagonists of

XX specific integrin receptors. The methods and the antagonists are useful

XX for inhibiting angiogenesis, tumour growth or metastases, or endothelial

XX cell interaction with the extracellular matrix. The antagonists are also

XX useful for treating diseases and conditions with accompanying undesired

XX angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,

XX carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,

XX neuroblastoma, osteosarcoma or leukaemia). These are also applicable to

XX treating non-tumorigenic diseases and conditions with accompanying

XX undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,

XX retinal neovascularisation, choroidal neovascularisation, muscular

XX degeneration, corneal graft rejection, vitamin A deficiency, atopic

XX keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,

XX sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser

XX complications, chronic inflammation or psoriasis.

Sequence 268 AA:

QY 5 KRQDGSPPATWTRGRFVETRHSSQTTAIPSCPECTVPLYSGFLEVOGNRAHQDGLT 64

DB 29 krgdgsppatwtrgrfvtthsgtlaipscpegtvplysgfslfvgngraqbgdgtl 88

QY 65 GSCLORETTMPFLFCNNVNDVCFASRNDYSYWLSTPALMNMNAPITGRALPEYISRCTV 124

DB 89 gsclorettmplfrcnnvndvcnfasrndysywlstpalmmnmapitgralepyisrcv 148

QY 125 CEGPAIAI 132

DB 149 cegpaiaai 156

Query Match 97.1%; Score 700; DB 21; Length 268;

Best Local Similarity 100.0%; Pred. No. 4.8e-70;

Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KRQDGSPPATWTRGRFVETRHSSQTTAIPSCPECTVPLYSGFLEVOGNRAHQDGLT 64

DB 29 krgdgsppatwtrgrfvtthsgtlaipscpegtvplysgfslfvgngraqbgdgtl 88

QY 65 GSCLORETTMPFLFCNNVNDVCFASRNDYSYWLSTPALMNMNAPITGRALPEYISRCTV 124

DB 89 gsclorettmplfrcnnvndvcnfasrndysywlstpalmmnmapitgralepyisrcv 148

QY 125 CEGPAIAI 132

DB 149 cegpaiaai 156

Search completed: March 6, 2002, 06:56:24

Job time: 135 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 06:55:22 ; Search time 26.9 Seconds  
(without alignments)  
110.425 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_1\_132

Perfect score: 721

Sequence: 1 GIKGRGDSGSPATWTTRGF.....RALEPPYISRCTVCEGPATAI 132

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

To number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%

Maximum Match 100%

Listing first 1000 summaries

Database :

Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Length	ID	Description
------------	-------------	--------	----	-------------

No matches found

Search completed: March 6, 2002, 06:55:22  
Job time: 73 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 06:54:49 ; Search time 31.16 Seconds

(without alignments)  
156,456 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_181\_244

Perfect score: 353

Sequence: 1 EFRAPELECHGRTGTCNYS.....KAGLEKTIISRCQVCMKKRH 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 219241 seqs, 76174552 residues

Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%

Database: PIR\_68:\*\*

1: PIR1:\*\*  
2: PIR2:\*\*  
3: PIR3:\*\*  
4: PIR4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	100.0	1670	1 CGHU3B	collagen alpha 3(I
2	345	97.7	246	2 I48302	collagen alpha 3(I
3	323	91.5	471	2 A39024	collagen alpha 3(I

#### ALIGNMENTS

#### RESULT 1

CGHU3B

N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form

C:Species: Homo sapiens (man)

C>Date: 28-Oct-1994 #sequence, revision 03-Oct-1995 #text\_change 22-Jun-1999

C:Accession: A54763; A43928; A44043; A45971; A39786

J:Biol. Chem. 269, 23013-23017, 1994

A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression

A:Reference number: A54763; MUID:94364994

A:Molecule type: mRNA

A:Residues: 1-1670 <MAP>

A:Cross-references: GB:X80031; NID:9577563; PID:9577564

A:Experimental source: kidney

R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.

J. Clin. Invest. 89, 592-601, 1992

A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the a

A:Reference number: A43928; MUID:92147878

A:Accession: A43928

A:Molecule type: mRNA

A:Residues: 1331-1524, '1', 1526-1670 <TUR>

A:Cross-references: GB:M81379

A:Experimental source: kidney

R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 267, 19780-19784, 1992

A:Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpas

tion.

A:Reference number: A44043; MUID:93015826

A:Accession: A44043

A:Molecule type: DNA: mRNA

A:Residues: 1386-1670 <OUT>

A:Cross-references: GB:M92993; NID:g177895; PIDN:AAA21610.1; PID:g177896

A:Note: sequence extracted from NCBI backbone (NCBI:115597)

R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 269, 17358, 1994

A:Reference number: A44738; MUID:94274734

A:Contents: annotation; erratum; correction to intronic sequence in A44043

R:Bernal, D.; Quinones, S.; Saus, J.

J. Biol. Chem. 268, 12090-12094, 1993

A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.

A:Reference number: A45971; MUID:93280184

A:Accession: A45971

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1427-1444 <BER>

A:Note: sequence extracted from NCBI backbone (NCBI:133363); sequence incorrectly id

R:Morrison, K.E.; Maruyama, M.; Yang-Feng, T.L.; Reeders, S.T.

Am. J. Hum. Genet. 49, 545-554, 1991

A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain

A:Reference number: A39786; MUID:91335370

A:Accession: A39786

A:Molecule type: mRNA

A:Residues: 1453-1593, 'A', 1595-1670 <MOR>

A:Cross-references: GB:S55790; NID:g234418; PIDN:AAB19637.1; PID:g234419

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit

ed and subsequently O-glycosylated.

C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epit

C:Genetics:

A:Gene: GDB:COL4A3

A:Cross-references: GDB:128351; OMIM:120070

A:Map position: 2q36-qq37

A:Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete

A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands w

C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha

monomer trimers amino-terminal domains (with disulfide and desmosine cross-links), dimeri

er associations in the interrupted helical domain (with disulfide and desmosine cross

C:Function:

A:Description: minor structural component of extracellular basement membrane in kidne

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extra

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MA

F:29-42/Domain: amino-terminal nonhelical, NHI <NHI>

F:43-1438/Region: Interrupted helical

F:791-793/Region: cell attachment (R-G-D) motif

F:996-998/Region: cell attachment (R-G-D) motif

F:1154-1156/Region: cell attachment (R-G-D) motif

F:1306-1308/Region: cell attachment (R-G-D) motif

F:1345-1347/Region: cell attachment (R-G-D) motif

F:1432-1434/Region: cell attachment (R-G-D) motif

F:1439-1440/Domain: carboxyl-terminal nonhelical, NCI <NCI>

F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTR1>

F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR2>

F:31, 33, 39, 41, 125, 422, 476, 479, 682, 722, 809, 1387/Disulfide bonds: interchain #status pr

F:253/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:1460-1548, 1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted

F:1505-1511, 1616-1622/Disulfide bonds: #status predicted

F:1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 100.0%; Score 353; DB 1; Length 1670;

Best Local Similarity 100.0%; Pred. No. 5, 5e-35;





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 07:05:09 ; Search time 18.75 Seconds  
(without alignments)  
125.149 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_181\_244  
Perfect score: 353  
Sequence: 1 EFRASPLECHGKCTCNWYS.....KAGELEKIISRCQYCMKKRH 64

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

To: number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%

Maximum Match 100%

Listing first 1000 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	353	100.0	1670	1 CA34_HUMAN	001955 homo sapien
2	323	91.5	471	1 CA34_BOVIN	028084 bos taurus

#### ALIGNMENTS

RESULT 1  
CA34\_HUMAN STANDARD: PRT; 1670 AA.  
AC 001955;  
DT 01-OCT-1996 (rel. 34, Created)  
DT 15-JUL-1999 (rel. 38, Last sequence update)  
DT 20-AUG-2001 (rel. 40, Last annotation update)  
DE COLLAGEN ALPHA 3(IV) CHAIN PRECURSOR (GOODPASTURE ANTIGEN).  
GN COL4A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RA MEDLINE=9436494; PubMed=8083201;  
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reeders S.T.;  
RT "Complete primary structure of the human alpha 3(IV) collagen chain.  
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  
RT human tissues.";  
RT J. Biol. Chem. 269:23013-23017(1994).  
RN (2)  
RP REVISIONS.  
RA Leinonen A.;  
RA Submitted (Oct-1998) to the EMBL/Genbank/DBJ databases.

RN (3)  
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=93015826; PubMed=1400291;  
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;  
RT "Exon/Intron structure of the human alpha 3(IV) gene encompassing the  
RT Goodpasture antigen (alpha 3(IV)NCL). Identification of a potentially  
RT antigenic region at the triple helix/NCL domain junction.";  
RT J. Biol. Chem. 267:19780-19784(1992).  
RN (4)  
RP SEQUENCE OF 1453-1670 FROM N.A.  
RX MEDLINE=91353570; PubMed=1882840;  
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reeders S.T.;  
RT "Sequence and localization of a partial cDNA encoding the human alpha  
RT 3 chain of type IV collagen.";  
RT Am. J. Hum. Genet. 49:545-554(1991).  
RN (5)  
RP SEQUENCE OF 1331-1670 FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=92147878; PubMed=1737849;  
RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
RA Pusey C.D.;  
RT "Molecular cloning of the human Goodpasture antigen demonstrates it  
RT to be the alpha 3 chain of type IV collagen.";  
RT J. Clin. Invest. 89:592-601(1992).  
RN (6)  
RP SEQUENCE OF 1644-1670 FROM N.A.  
RC TISSUE=Kidney;  
RA Ding J.;  
RT Submitted (JAN-1993) to the EMBL/Genbank/DBJ databases.  
RN (7)  
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.  
RC TISSUE=Kidney;  
RX MEDLINE=94124597; PubMed=8294492;  
RA Feng L., Xia Y., Wilson C.B.;  
RT "Alternative splicing of the NCL domain of the human alpha 3(IV)  
RT collagen gene. Differential expression of mRNA transcripts that  
RT predict three protein variants with distinct carboxyl regions.";  
RT J. Biol. Chem. 269:2342-2348(1994).  
RN (8)  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=98196854; PubMed=9537506;  
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,  
RA Nimomiya Y.;  
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
RT 2q35.";  
RT FEBS Lett. 424:11-16(1998).  
RN (9)  
RP ALTERNATIVE SPLICING.  
RX MEDLINE=93280184; PubMed=8505332;  
RA Bernal D., Quinones S., Saus J.;  
RT "The human mRNA encoding the Goodpasture antigen is alternatively  
RT spliced.";  
RT J. Biol. Chem. 268:12090-12094(1993).  
RN (10)  
RP VARIANT PRO-1474.  
RX MEDLINE=95078827; PubMed=7987301;  
RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
RA Bartelentos A., Monnens L.A.H., van Oost B.A., Brunner H.G.;  
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
RT recessive Alport syndrome.";  
RT Hum. Mol. Genet. 3:1269-1273(1994).  
RN (11)  
RP VARIANTS AS, AND VARIANTS.  
RX MEDLINE=21064696; PubMed=11134255;  
RA Heidet L., Arondel C., Forestier L., Cohen-Solal L., Mollet G.,  
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
RT autosomal Alport syndrome.";  
RT J. Am. Nephrol. 12:97-106(2001).  
CC - FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'

FT	SITE	791	793	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	966	998	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1154	1156	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1306	1308	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1345	1347	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1432	1434	CELL ATTACHMENT SITE (POTENTIAL).
FT	CAROHXD	253	253	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	MOD_RES	1435	1435	PHOSPHORYLATION (BY SIMILARITY).
FT	MOD_RES	1437	1437	PHOSPHORYLATION (BY SIMILARITY).
FT	DISULFID	1460	1451	OR 1548 (BY SIMILARITY).
FT	DISULFID	1493	1548	OR 1551 (BY SIMILARITY).
FT	DISULFID	1505	1511	BY SIMILARITY.
FT	DISULFID	1570	1665	OR 1662 (BY SIMILARITY).
FT	DISULFID	1604	1662	OR 1665 (BY SIMILARITY).
FT	DISULFID	1616	1622	BY SIMILARITY.
FT	VARSPLIC	1586	1670	FTSASEGGOALASPGSCLEBFRRASPFLCHGRCNTNYS
FT				NSYSFWLASINPERMFRKPIPTVYAGLELEKIISRCVCM
FT	VARSPLIC	1488	1670	KRH -> KAYSINCESWGIRKNKNSLSGVHEKTLKKTA
FT				ELVFFILKKVMEHAVI (IN ISOFORM 2).
FT				GLFGSCLOFRTTMEPLFCQVNDVCPNPSANDSVYLSPTAL
FT				MPMANAPITGRALPEYISRCYCECPALAIYAVHSOTDIP
FT				CPHGNISLWKGSEFTIMFTSAGEGGOALASGSCLEEFRA
FT				SPFLCHGRCNTNYSNSYSFWLASINPERMFRKPIPTVYK
FT				AGELEKIISRCVCMKRRH -> DALFVAVLRLSP (IN
FT				ISOFORM 3).
FT				G -> R.
FT	VARIANT	43	43	/FTID=VAR_011202.
FT				G -> E.
FT	VARIANT	162	162	/FTID=VAR_011203.
FT				G -> E (IN AS).
FT	VARIANT	297	297	
Query Match			100.0%;	Score 353; DB 1; Length 1670;
Best Local Similarity			100.0%;	Pred. No. 2.7e-36;
Matches 64;	Conservative		0;	Mismatches 0; Indels 0; Gaps 0
OY	1	ERRASPFLCHGRCNTNYSNSYSFWLASINPERMFRKPIPTVYAGLELEKIISRCVCM	60	
DB	1607	ERRASPFLCHGRCNTNYSNSYSFWLASINPERMFRKPIPTVYAGLELEKIISRCVCM	1666	
OY	61	KRRH	64	
DB	1667	KRRH	1670	
RESULT	2			
CA34_BOVIN		STANDARD:	PRT:	471 AA.
ID	CA34_BOVIN			
AC	028084:			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	COLLAGEN ALPHA 3(IV) CHAIN (FRAGMENT).			
GN	COL4A3.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lens;			
RC	MEDLINE=91093146; PubMed=1985905;			
RA	Morrison K.E., Germino G.G., Reeders S.T.;			
RT	"Use of the polymerase chain reaction to clone and sequence a cDNA			
RT	encoding the bovine alpha 3 chain of type IV collagen.";			
RL	J. Biol. Chem. 266:34-39(1991).			
RN	[2]			
RP	SEQUENCE OF 227-258.			
RC	TISSUE=Kidney;			
RC	MEDLINE=90202779; PubMed=2318622;			
RA	Gunwar S., Saus J., Noelken M.E., Hudson B.G.;			
RT	"Glomerular basement membrane. Identification of a fourth chain,			
RT	alpha 4, of type IV collagen.";			

RL J. Biol. Chem. 265:5466-5469(1990).  
 RP SEQUENCE OF 227-254.  
 RA MEDLINE-88330844; Pubmed-3417661;  
 RA Saus J., Wieslander J.P.M., Quinones S., Hudson B.G.;  
 RT "Identification of the goodpasture antigen as the alpha 3(IV) chain  
 of collagen IV.";  
 RT J. Biol. Chem. 263:13374-13380(1988).  
 RN [4]  
 RP SEQUENCE OF 227-244.  
 RX MEDLINE-87222419; Pubmed-2438283;  
 RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,  
 RA Hudson B.G.;  
 RT "Localization of the Goodpasture epitope to a novel chain of basement  
 membrane collagen.";  
 RT J. Biol. Chem. 262:7874-7877(1987).  
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENACTIN/  
 NIDOGEN.  
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-  
 ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 IV COLLAGENS.  
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M63139; AAA62708.1; -  
 DR InterPro: IPR001442; C4;  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01413; C4; 2.  
 DR ProDom: PD003923; C4; 2.  
 DR SMART: SM00111; C4; 2.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
 FT NON\_TER 1 1  
 FT DOMAIN <1 238 TRIPLE-HELICAL REGION.  
 FT DOMAIN 239 471 NONHELICAL REGION (NC1).  
 FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).  
 FT MOD\_RES 232 232 HYDROXYLATION.  
 FT MOD\_RES 238 238 HYDROXYLATION.  
 FT DISULFID 261 352 OR 349 (BY SIMILARITY).  
 FT DISULFID 294 349 OR 352 (BY SIMILARITY).  
 FT DISULFID 306 312 BY SIMILARITY.  
 FT DISULFID 371 466 OR 463 (BY SIMILARITY).  
 FT DISULFID 405 463 OR 466 (BY SIMILARITY).  
 FT DISULFID 417 423 BY SIMILARITY.  
 FT CONFLICT 253 253 S -> Y (IN REF. 3).  
 SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E708DE CRC64;

OY	1	EFRAPELECHGRCYCNYSNSYSFWLASLNPERRKRPISPTVKAGELKIIISRCQVM	60
Db	408	EFRAPELECHGRCYCNYSNSYSFWLASLDPKRMFRKPIPTVKAGELNIISRCQVM	467
OY	61	KKR	63
Db	468	KMR	470

Search completed: March 6, 2002, 07:05:09  
 Job time: 585 sec

Query Match 91.5%; Score 323; DB 1; Length 471;  
 Best Local Similarity 92.1%; Pred. No. 4e-33;  
 Matches 58; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 07:04:44 ; Search time 50.56 Seconds  
(without alignments)  
185.155 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_181\_244  
Perfect score: 353  
Sequence: 1 EFRASPFLCHGRTGCTCNVYS.....KAGELEKTIISRCQVCMKKRH 64

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
T number of hits satisfying chosen parameters: 7

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%  
Maximum Match 100%  
Listing first 1000 summaries

Database :  
1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_RODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	100.0	212	6	Q28512 macaca mula
2	353	100.0	245	4	Q9NYC4 O9NYC4 homo sapien
3	353	100.0	1670	4	Q9BQT2 O9bqt2 homo sapien
4	348	98.6	230	11	Q63122 ratulus nov
5	348	98.6	246	11	Q61435 mus musculu
6	348	98.6	1669	11	Q9Q2S0 O9q2s0 mus musculu
7	331	93.8	212	6	Q28567 ovis aries

## ALIGNMENTS

RESULT 1  
ID Q28512 PRELIMINARY; PRT: 212 AA.  
AC Q28512;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).  
GN COL4A3.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=RENAL CORTEX;  
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbama I.,  
RA Mason P.J., Pusey C.D.;  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: I47280; AAA91861.1; -  
DR InterPro: IPR000504; RRM.  
DR InterPro: IPR001442; C4.  
DR Pfam: PF01413; C4; 2.  
DR ProDom: PD003923; C4; 2.  
DR PROSITE: PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
DR SMART: SM00111; C4; 2.  
FT NON\_TER 1  
FT NON\_TER 212  
SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357B64 CRC64;

Query Match 100.0%; Score 353; DB 6; Length 212;  
Best Local Similarity 100.0%; Pred. No. 1.8e-36;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EFRASPFLCHGRTGCTCNVYSFWMASLNPERFRKPIPTVAKAGELEKTIISRCQVCM 60  
Db 149 EFRASPFLCHGRTGCTCNVYSFWMASLNPERFRKPIPTVAKAGELEKTIISRCQVCM 208  
QY 61 KKRH 64  
Db 209 KKRH 212

RESULT 2  
ID Q9NYC4 PRELIMINARY; PRT: 245 AA.  
AC Q9NYC4;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE TMSSTATTIN (FRAGMENT).  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,  
RA Erickson M.D., Hopfer H., Xiao Y., Stillman I.E., Kallunki R.;  
RT "Distinct anti-tumor properties of a type IV collagen domain derived from basement membrane."  
RL J. Biol. Chem. 0:0-0(2000).  
DR EMBL: AF258351; AAF72632.1; -  
DR InterPro: IPR001442; C4.  
DR InterPro: IPR000504; RRM.  
DR Pfam: PF01413; C4; 2.  
DR ProDom: PD003923; C4; 2.  
DR SMART: SM00111; C4; 2.  
DR PROSITE: PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 245 AA; 26952 MW; 1EE5028354D9A57D CRC64;

Query Match 100.0%; Score 353; DB 4; Length 245;  
Best Local Similarity 100.0%; Pred. No. 2.1e-36;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EFRASPFLCHGRTGCTCNVYSFWMASLNPERFRKPIPTVAKAGELEKTIISRCQVCM 60

```

DB      182 EFRA5PFLCHGRGTCNYSNSYSFWLASLNPERMFRKPIPTVKAGLEKITSRCQVCM 241
QY      61 KKRH 64
DB      242 KKRH 245

RESULT 3
09BOT2 PRELIMINARY; PRT; 1670 AA.
AC 09BOT2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DE ALPNA3 TYPE IV COLLAGEN.
GN COLA43.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
[1]
RX MEDLINE=21064696; PubMed=11134255;
RA Heider L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT "Structure of the human type IV collagen gene COLA43 and mutations in
RT autosomal Alport syndrome."
RL J. Am. Soc. Nephrol. 12:97-106(2001).
DR EMBL; AJ288487; CAC36101.1; JOINED.
DR EMBL; AJ288488; CAC36101.1; JOINED.
DR EMBL; AJ288489; CAC36101.1; JOINED.
DR EMBL; AJ288490; CAC36101.1; JOINED.
DR EMBL; AJ288491; CAC36101.1; JOINED.
DR EMBL; AJ288492; CAC36101.1; JOINED.
DR EMBL; AJ288493; CAC36101.1; JOINED.
DR EMBL; AJ288494; CAC36101.1; JOINED.
DR EMBL; AJ288495; CAC36101.1; JOINED.
DR EMBL; AJ288496; CAC36101.1; JOINED.
DR EMBL; AJ288497; CAC36101.1; JOINED.
DR EMBL; AJ288498; CAC36101.1; JOINED.
DR EMBL; AJ288499; CAC36101.1; JOINED.
DR EMBL; AJ288500; CAC36101.1; JOINED.
DR EMBL; AJ288501; CAC36101.1; JOINED.
DR EMBL; AJ288502; CAC36101.1; JOINED.
DR EMBL; AJ288503; CAC36101.1; JOINED.
DR EMBL; AJ288504; CAC36101.1; JOINED.
DR EMBL; AJ288505; CAC36101.1; JOINED.
DR EMBL; AJ288506; CAC36101.1; JOINED.
DR EMBL; AJ288507; CAC36101.1; JOINED.
DR EMBL; AJ288508; CAC36101.1; JOINED.
DR EMBL; AJ288509; CAC36101.1; JOINED.
DR EMBL; AJ288510; CAC36101.1; JOINED.
DR EMBL; AJ288511; CAC36101.1; JOINED.
DR EMBL; AJ288512; CAC36101.1; JOINED.
DR EMBL; AJ288513; CAC36101.1; JOINED.
DR EMBL; AJ288514; CAC36101.1; JOINED.
DR EMBL; AJ288515; CAC36101.1; JOINED.
DR EMBL; AJ288516; CAC36101.1; JOINED.
DR EMBL; AJ288517; CAC36101.1; JOINED.
DR EMBL; AJ288518; CAC36101.1; JOINED.
DR EMBL; AJ288519; CAC36101.1; JOINED.
DR EMBL; AJ288520; CAC36101.1; JOINED.
DR EMBL; AJ288521; CAC36101.1; JOINED.
DR EMBL; AJ288522; CAC36101.1; JOINED.
DR EMBL; AJ288523; CAC36101.1; JOINED.
DR EMBL; AJ288524; CAC36101.1; JOINED.
DR EMBL; AJ288525; CAC36101.1; JOINED.
DR EMBL; AJ288526; CAC36101.1; JOINED.
DR EMBL; AJ288527; CAC36101.1; JOINED.
DR EMBL; AJ288528; CAC36101.1; JOINED.
DR EMBL; AJ288529; CAC36101.1; JOINED.
DR EMBL; AJ288530; CAC36101.1; JOINED.

```

```

DR EMBL; AJ288531; CAC36101.1; JOINED.
DR EMBL; AJ288532; CAC36101.1; JOINED.
DR EMBL; AJ288533; CAC36101.1; JOINED.
DR EMBL; AJ288534; CAC36101.1; JOINED.
DR EMBL; AJ288535; CAC36101.1; JOINED.
DR EMBL; AJ288536; CAC36101.1; JOINED.
DR EMBL; AJ288537; CAC36101.1; JOINED.
DR EMBL; AJ288538; CAC36101.1; JOINED.
KW Collagen.
SQ SEQUENCE 1670 AA; 161899 MW; FA7BE4914CA0A6F6 CRC64;

```

```

Query Match 100.0%; Score 353; DB 4; Length 1670;
Best Local Similarity 100.0%; Pred. No. 1,6e-35;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 EFRA5PFLCHGRGTCNYSNSYSFWLASLNPERMFRKPIPTVKAGLEKITSRCQVCM 60
DB 1607 EFRA5PFLCHGRGTCNYSNSYSFWLASLNPERMFRKPIPTVKAGLEKITSRCQVCM 1666
QY 61 KKRH 64
DB 1667 KKRH 1670

```

```

RESULT 4
063122 PRELIMINARY; PRT; 230 AA.
AC 063122;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE ALPNA-3 TYPE IV COLLAGEN (FRAGMENT).
GN COLA43.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
RX SEQUENCE FROM N.A.
RX STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY CORTEX;
RA Turner N., Ryan J.J., Derry C.J., Cashman S.J., Katbama I.,
RA Mason P.J., Pusey C.D.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; I47281; AAB72238.2;
DR InterPro; IPR005054; RKM.
DR InterPro; IPR001442; C4.
DR Pfam; PF01413; C4; 2.
DR ProDom; PD003923; C4; 2.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
DR SMART; SM00111; C4; 2.
FT NON_TER 1
FT NON_TER 230
SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

```

```

Query Match 98.6%; Score 348; DB 11; Length 230;
Best Local Similarity 96.9%; Pred. No. 8,5e-36;
Matches 62; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 EFRA5PFLCHGRGTCNYSNSYSFWLASLNPERMFRKPIPTVKAGLEKITSRCQVCM 60
DB 167 EFRA5PFLCHGRGTCNYSNSYSFWLASLNPERMFRKPIPTVKAGLEKITSRCQVCM 226
QY 61 KKRH 64
DB 227 KKRH 230

```

```

RESULT 5
061435 PRELIMINARY; PRT; 246 AA.
AC 061435;

```

DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE PROCOLLAGEN TYPE IV, ALPHA 3 (COLLAGEN IV ALPHA 3 CHAIN) (FRAGMENT).  
 GN COL4A3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RX MEDLINE=95050957; PubMed=7962065;  
 RA Miner J.H., Janes J.R.;  
 RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal  
 laminae: sequence, distribution, association with laminins, and  
 developmental switches."  
 RL J. Cell Biol. 127:879-891(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RA Miner J.H.;  
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C;  
 RA Miner J.H.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z35156; CAA84529.1; -.  
 DR MGD: MGI:104688; Col4a3.  
 DR InterPro: IPR001442; C4.  
 DR InterPro: IPR000504; RRM.  
 DR Pfam: PF01413; C4; 2.  
 DR ProDom: PD003923; C4; 2.  
 DR SMART: SM00111; C4; 2.  
 DR PROSITE: PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 FT NON\_TER 1  
 SQ SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;  
 Query Match 98.6%; Score 348; DB 11; Length 246;  
 Best Local Similarity 96.9%; Pred. No. 9.1e-36;  
 Matches 62; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EFRRAPFLECHGRCCTCNVNSYSFWMASLNPERMFRKPIPTVKAGLEKTIISRCQVCM 60  
 DB 183 EFRRAPFLECHGRCCTCNVNSYSFWMASLNPERMFRKPIPTVKAGLEKTIISRCQVCM 242  
 O 61 KKRH 64  
 DB 243 KKRH 246  
 RESULT 6  
 Q09Q250 PRELIMINARY; PRT; 1669 AA.  
 AC Q09Q250;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE ALPHA 3 COLLAGEN IV.  
 GN COL4A3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-KIDNEY;  
 RX MEDLINE=20005934; PubMed=10534397;  
 RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,  
 RA Elder F.F.B., Miner J.H., Overbeek P.A., Meister M.H.;  
 RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a

RT mouse model of alport syndrome.";  
 RL Genomics 61:113-124(1999).  
 DR EMBL: AF169387; AAD50449.1; -.  
 DR MGD: MGI:104688; Col4a3.  
 DR InterPro: IPR001442; C4.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000504; RRM.  
 DR Pfam: PF01413; C4; 2.  
 DR Pfam: PF01391; Collagen; 21.  
 DR ProDom: PD003923; C4; 2.  
 DR SMART: SM00111; C4; 2.  
 DR PROSITE: PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 FT COLLAGEN.  
 SQ SEQUENCE 1669 AA; 161769 MW; 30976E59739A47B2 CRC64;  
 Query Match 98.6%; Score 348; DB 11; Length 1669;  
 Best Local Similarity 96.9%; Pred. No. 6.9e-35;  
 Matches 62; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EFRRAPFLECHGRCCTCNVNSYSFWMASLNPERMFRKPIPTVKAGLEKTIISRCQVCM 60  
 DB 1606 EFRRAPFLECHGRCCTCNVNSYSFWMASLNPERMFRKPIPTVKAGLEKTIISRCQVCM 1665  
 QY 61 KKRH 64  
 DB 1666 KKRH 1669  
 RESULT 7  
 Q28567 PRELIMINARY; PRT; 212 AA.  
 AC Q28567;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).  
 GN COL4A3.  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-RENAL CORTEX;  
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbama I.,  
 RA Mason P.J., Pusey C.D.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: I47282; AAA91904.1; -.  
 DR InterPro: IPR000504; RRM.  
 DR InterPro: IPR001442; C4.  
 DR Pfam: PF01413; C4; 2.  
 DR ProDom: PD003923; C4; 2.  
 DR PROSITE: PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 DR SMART: SM00111; C4; 2.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDB8C CRC64;  
 Query Match 93.8%; Score 331; DB 6; Length 212;  
 Best Local Similarity 92.1%; Pred. No. 1.1e-33;  
 Matches 58; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EFRRAPFLECHGRCCTCNVNSYSFWMASLNPERMFRKPIPTVKAGLEKTIISRCQVCM 60  
 DB 149 EFRRAPFLECHGRCCTCNVNSYSFWMASLNPERMFRKPIPTVKAGLEKTIISRCQVCM 208  
 QY 61 KKR 63  
 DB 209 KKR 211

Wed Mar 6 07:07:42 2002

Search completed: March 6, 2002, 07:04:44  
Job time: 590 sec

us-09-543-371-10\_copy\_181\_244.rspt



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 06:56:24 ; Search time 54.67 Seconds  
(without alignments)  
86,715 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_181\_244

Perfect score: 353  
Sequence: 1 EFRASPLECHGRCGTCTNYNS.....KAGELEKTIISRCQVCMKRH 64

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Number of hits satisfying chosen parameters: 9

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%

Maximum Match 100%

Listing first 1000 summaries

Database : A\_Geneseq\_1101.\*

1:	/SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	353	100.0	218	20	AAV44172 Human type IV coll
2	353	100.0	218	21	AAV56784 Human alpha3 type
3	353	100.0	245	21	AAV67942 Human type IV coll
4	353	100.0	268	20	AAV31993 Type IV collagen N
5	353	100.0	268	21	AAV97555 Human alpha3(IV)NC
6	358	95.8	218	16	AAV79164 Partial sequence o
7	323	91.5	471	16	AAV79163 Bovine type IV col
8	323	91.5	471	20	AAV44171 Bovine alpha3 type
9	323	91.5	471	21	AAV56783

## ALIGNMENTS

RESULT 1  
ID AAV44172 standard; Protein; 218 AA.  
XX  
AC AAV44172;  
XX  
DT 01-FEB-2000 (first entry)  
XX  
DE Human type IV collagen alpha3 chain protein.  
XX  
KW Recombinant; bovine; alpha3 chain; type IV collagen; detection;  
XX Goodpasture syndrome; antibody; blood; tissue; human; nephrotrophism.  
XX  
OS Homo sapiens.  
XX  
PN US5973120-A.  
XX  
PD 26-OCT-1999.  
XX  
PF 07-MAR-1995; 95US-0399889.  
XX  
PR 30-NOV-1990; 90US-0621091.  
XX  
PA (UYVA ) UNIV YALE.  
XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
XX  
PI Hudson BG, Reiders ST, Morrison KE;  
XX WPI; 1999-610317/52.  
XX DR N-PSDB; AA228775.  
XX  
PT Isolated alpha 3 chain of type IV collagen polypeptide useful for  
XX diagnosis and treatment of Goodpasture syndrome -  
XX  
PS Claim 2; Column 35-36; 27pp; English.  
XX  
CC This sequence represents a recombinant human alpha3 chain of type IV  
XX collagen polypeptide. The sequence corresponds to the 218 amino acids  
XX of the C-terminal non-collagenous domain. Alpha3 chain collagen  
XX polypeptides are useful for detecting Goodpasture antibodies in blood  
XX or tissue from a human patient and for treating Goodpasture syndrome,  
XX especially by neutralizing the antibodies in the blood. The polypeptides  
XX also have a nephrotrophic activity.  
SQ Sequence 218 AA:  
Query Match 100.0%; Score 353; DB 20; Length 218;  
Best Local Similarity 100.0%; Pred. No. 9e-39;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EFRASPLECHGRCGTCTNYNSYSFWLASINPERFRKPIPTVKAGELEKTIISRCQVM 60  
DB 155 EFRASPLECHGRCGTCTNYNSYSFWLASINPERFRKPIPTVKAGELEKTIISRCQVM 214  
QY 61 KKRH 64  
DB 215 KKRH 218  
RESULT 2  
ID AAV56784 standard; Protein; 218 AA.  
XX  
AC AAV56784;  
XX  
DT 27-MAR-2000 (first entry)  
XX  
DE Human alpha3 type IV collagen C-terminal domain.  
XX

XX	17-JUN-1999;	99WO-US13737.
XX	17-JUN-1998;	98US-0089689.
PR	25-MAR-1999;	99US-0126175.
XX		
PA	(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.	
XX		
PI	Kalluri R;	
DR	WPI; 2000-097708/08.	
DR	N-PSDB; AA257158.	
XX		
PT	Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or	
PT	3 chain of type IV collagen used in, e.g. treatment of benign tumors	
PT	and rheumatoid arthritis -	
XX		
PS	Claim 32; Flg 16B; 117pp; English.	
XX		
CC	The present sequence represents the human type IV collagen alpha 3 chain.	
CC	The present invention describes an isolated protein chosen from the NC1	
CC	domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or	
CC	a fragment, analogue, derivative or mutant, which has anti-angiogenic	
CC	properties. The anti-angiogenic proteins, multimers and chimeras are	
CC	useful for inhibiting angiogenic activity in mammalian tissue,	
CC	especially for treating diseases chosen from angiogenesis-dependent	
CC	cancers, benign tumours, rheumatoid arthritis, diabetic retinopathy,	
CC	psoriasis, ocular angiogenesis diseases, Osler-Weber Syndrome,	
CC	myocardial angiogenesis, plaque neovascularisation, telangiectasia,	
CC	hemophilic joints, angiolipoma, wound granulation, intestinal	
CC	adhesions, atherosclerosis, scleroderma, hypertrophic scars, cat scratch	
CC	disease, Helicobacter pylori ulcers, dialysis graft vascular access	
CC	stenosis, contriception and obesity. The compositions can be used to	
CC	inhibit a disease characterised by angiogenic activity, in conjunction	
CC	with radiation therapy, chemotherapy or immunotherapy.	
XX		
SO	Sequence 245 AA;	
	Query Match	100.0%; Score 353; DB 21; Length 245;
	Best Local Similarity 100.0%; Pred. No. 1e-38;	
	Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 EPRASPLECHGKGTCTNTYSNSYSFWLASLNDERMKRPDPSTVAGLEKITSQCYCM 60	
DB	182 efRASpLEChgRgycnYnsysfWlaslndErMfRxpDpStvAgelEkITSrCYcm 241	
QY	61 KKRH 64	
DB	242 Kkrh 245	
	RESULT 4	
AA131993	AA131993 standard; Protein: 268 AA.	
XX		
AC	AA131993;	
XX		
DC	05-JAN-2000 (first entry)	
XX		
DE	Type IV collagen NC1 domain alpha-3 monomer.	
KW	Type IV collagen; NC1 domain; non-collagenous domain; human;	
KW	angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;	
KW	rheumatoid arthritis; retinal neovascularization;	
KW	choroidal neovascularization; macular degeneration;	
KW	corneal neovascularization; retinopathy of prematurity;	
KW	corneal graft rejection; neovascular glaucoma;	
KW	retrolental fibroplasia; epidemic keratoconjunctivitis;	
KW	vitamin A deficiency; contact lens overwear; atopic keratitis;	
KW	superior limbic keratitis; pterygium keratitis sicca; sores;	
KW	acne rosacea; phlyctenulosis; syphilis; Mycobacteria infection;	
KW	lipid degeneration; chemical burn; ulcer; herpes simplex infection;	
KW	Herpes zoster infection; protozoan infection; Kaposi's sarcoma;	

KW Mooren ulcer; Terrien's marginal degeneration;  
 KW marginal keratolysis; trauma; systemic lupus; polyarteritis;  
 KW Wegener's sarcoidosis; scleritis; Steven's Johnson disease;  
 KW radial keratotomy; sickle cell anaemia; sarcoid;  
 KW pseudoxanthoma elasticum; Paget's disease; vein occlusion;  
 KW artery occlusion; carotid obstructive disease; chronic uveitis;  
 KW chronic vitritis; Lyme's disease; Eales disease; Behcet's disease;  
 KW myopia; optic pit; Stargardt's disease; pars planitis;  
 KW chronic retinal detachment; hyperviscosity syndrome; toxoplasmosis;  
 KW post-laser complication; fibrovascular tissue proliferation;  
 KW haemangioma; Osler-Weber-Rendu; AIDS; ocular neovascular disease;  
 KW osteoarthritis; chronic inflammation; Crohn's disease;  
 KW ulcerative colitis; psoriasis; atherosclerosis; pemphigoid.  
 XX Homo sapiens.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..17  
 FT /note= "BM40 signal peptide"  
 FT Protein 18..268  
 FT /note= "mature protein"  
 FT Peptide 18..25  
 FT /note= "affinity tag"  
 FT Protein 26..268  
 FT /note= "NC1 alpha-3 monomer"  
 XX  
 XX MO9949885-AZ.  
 XX  
 XX 07-OCT-1999.  
 XX  
 XX 26-MAR-1999; 99MO-US06445.  
 XX  
 XX 27-MAR-1998; 98US-0079783.  
 XX 29-OCT-1998; 98US-0106170.  
 XX  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 XX  
 XX Hudson BG, Sarraf MP;  
 XX  
 XX MPI: 1999-601297/51.  
 XX N-PSDB; AA220091.  
 XX  
 XX Inhibition of angiogenesis with non-collagenous alpha chain monomer  
 XX useful for treating e.g. tumor growth or metastasis,  
 XX neovascularisation, etc -  
 XX  
 XX Disclosure: Fig 17c; 56pp; English.  
 XX  
 XX This sequence represents a recombinant type IV collagen  
 XX non-collagenous (NC1) domain alpha-3 polypeptide composed of a  
 XX BM40 signal sequence (which is cleaved from the mature protein) to  
 XX facilitate protein secretion, and a mature protein comprising an  
 XX affinity tag (facilitates purification and identification of the  
 XX material) and the alpha-1 chain monomer. The invention provides  
 XX methods and kits for inhibiting angiogenesis, tumour growth and  
 XX metastasis, and endothelial cell interaction with the extracellular  
 XX matrix, each method comprising contacting the tumour or animal  
 XX tissue with 1 or more isolated type IV collagen NC1 alpha chain  
 XX monomer(s) selected from the group consisting of alpha-1, alpha-2,  
 XX alpha-3 and alpha-6 NC1 chain monomers (see AAY31991-96). The  
 XX monomers can be produced via recombinant protein expression. The  
 XX polynucleotides and polypeptides are used to treat an angiogenesis-  
 XX mediated disorder or condition, especially selected from solid and  
 XX blood-borne tumours, diabetic retinopathy, rheumatoid arthritis,  
 XX retinal neovascularization, choroidal neovascularization, macular  
 XX degeneration, corneal neovascularization, retinopathy of prematurity,  
 XX corneal graft rejection, neovascular glaucoma, retrolental  
 XX fibroplasia, epidemic keratoconjunctivitis, vitamin A deficiency,  
 XX contact lens overwear, atopic keratitis, superior limbic keratitis,  
 XX pterygium keratitis sicca, seborrheic acne rosacea, phlyctenulosis,  
 XX syphilis, mycobacteria infections, lipid degeneration, chemical  
 XX burns, bacterial ulcers, fungal ulcers, herpes simplex infections,

CC herpes zoster infections, protozoan infections, Kaposi's sarcoma,  
 CC Mooren ulcer, Terrien's marginal degeneration, marginal keratolysis,  
 CC trauma, systemic lupus, polyarteritis, Wegener's sarcoidosis,  
 CC scleritis, Steven's Johnson disease, radial keratotomy, sickle cell  
 CC anaemia, sarcoid, pseudoxanthoma elasticum, Paget's disease, vein  
 CC occlusion, artery occlusion, carotid obstructive disease, chronic  
 CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Behcet's  
 CC disease, myopia, optic pits, Stargardt's disease, pars planitis,  
 CC chronic retinal detachment, hyperviscosity syndrome, toxoplasmosis,  
 CC post-laser complications, abnormal proliferation of fibrovascular  
 CC tissue, haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular  
 CC disease, osteoarthritis, chronic inflammation, Crohn's disease,  
 CC ulcerative colitis, psoriasis, atherosclerosis, and pemphigoid (all  
 CC claimed).  
 CC  
 XX  
 XX Sequence 268 AA:  
 SQ  
 Query Match 100.0%; Score 353; DB 20; Length 268;  
 Best Local Similarity 100.0%; Pred. No. 1..1e-38;  
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EFRASPTLECHGRTGCTNYNSYSFNLASINPERMPKRPISPVKAGELEKIIISRCQYCM 60  
 Db /205 etrasptlechgtrgctnynsysfwlasinpermrkripstcvkagelkistrcqcm 264  
 QY 61 KRRH 64  
 Db 265 krrh 268  
 DB  
 RESULT 5  
 AAY97555  
 ID AAY97555 standard; Protein; 268 AA.  
 XX  
 XX AAY97555;  
 AC  
 XX  
 XX 12-FEB-2001 (first entry)  
 DT  
 XX  
 XX Human alpha3(IV)NC1 protein sequence.  
 DE  
 XX  
 XX Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;  
 XX tumour growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;  
 KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;  
 KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;  
 KW muscular degeneration; corneal graft rejection; vitamin A deficiency;  
 KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;  
 KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;  
 KW chronic inflammation; psoriasis; therapy; alpha3(IV)NC1.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 XX MO200059532-A1.  
 PN  
 XX  
 XX 12-OCT-2000.  
 PD  
 XX  
 XX 31-MAR-2000; 2000MO-US08678.  
 PF  
 XX  
 XX 01-APR-1999; 99US-0127391.  
 PR  
 XX  
 XX (BIOS-) BIOSTRATUM INC.  
 PA  
 XX  
 XX Brooks P, Hudson B;  
 PI  
 XX  
 XX MPI: 2000-664962/64.  
 DR N-PSDB; AAA90993.  
 DR  
 XX  
 XX Use of antagonists of specific integrin receptors for inhibiting  
 PT angiogenesis, tumour growth or metastases, or endothelial cell  
 PT interactions with the extracellular matrix -  
 XX  
 XX Disclosure: Fig 17c; 78pp; English.  
 XX

CC This sequence is a human type IV collagen alpha chain monomer.  
 CC designated alpha3(IV)NC1. The invention relates to a method for  
 CC inhibiting angiogenesis, tumour growth or metastases, or endothelial cell  
 CC interactions with the extracellular matrix, comprising contacting the  
 CC cells or tissue with a polypeptide composition containing antagonists of  
 CC specific integrin receptors. The methods and the antagonists are useful  
 CC for inhibiting angiogenesis, tumour growth or metastases, or endothelial  
 CC cell interaction with the extracellular matrix. The antagonists are also  
 CC useful for treating diseases and conditions with accompanying undesired  
 CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,  
 CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,  
 CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to  
 CC treating non-tumorigenic diseases and conditions with accompanying  
 CC undesired angiogenesis, e.g., diabetic retinopathy, rheumatoid arthritis,  
 CC retinal neovascularisation, choroidal neovascularisation, muscular  
 CC degeneration, corneal graft rejection, vitamin A deficiency, atopic  
 CC keratitis. Mycobacteria infections, chemical burns, Kaposi's sarcoma,  
 CC sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser  
 CC complications, chronic inflammation or psoriasis.

Sequence 268 AA;

Query Match 100.0%; Score 353; DB 21; Length 268;  
 Best Local Similarity 100.0%; Pred. NO. 1.1e-38;  
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EFRRSPFLECHGRGTCNTYNSYSFWMASLNPERMFRKPIPTVKAGLEKITSRCQVCM 60  
 DB 205 efrrspflechgrgtrcnynsysfwlaslnpermrfrkpiptvkaglekitsrcqvc 264  
 OY 61 KKRH 64  
 ||||  
 Db 265 kkrh 268

RESULT 6  
 AAR79164  
 ID AAR79164 standard; Protein: 218 AA.  
 AC AAR79164;  
 DT 22-DEC-1995 (first entry)  
 DE Partial sequence of human alpha 3 chain of type IV collagen.  
 XX  
 KW Type IV collagen; alpha 3 chain; Alport syndrome; COL4A3 gene.  
 Homo sapiens.

PN US5424408-A.  
 XX 13-JUN-1995.  
 PD  
 XX 30-NOV-1990; 90US-0621091.  
 PF  
 XX 30-NOV-1990; 90US-0621091.  
 PR  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (UYA ) UNIV YALE.  
 XX  
 PI Hudson BG, Morrison KE, Reeders ST;  
 XX WPI: 1995-262631/34.  
 DR N-PDB; AAQ96291.  
 XX

PT CDNAs encoding human or bovine alpha-3 type 4 collagen peptide(s) -  
 PT useful for detection and therapeutic removal of antibodies associated  
 PT with Goodpasture syndrome  
 PS Disclosure; Columns 7-10; 33pp; English.  
 XX  
 CC Using the PCR with primers derived from each end of the known 27 AA

CC residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp.  
 CC bovine genomic fragment was amplified. This fragment was then used  
 CC to a bovine lens CDNA library and a 1.5 kb partial cDNA clone was  
 CC obt'd. (clone KMC15). This encodes 238 residues of the triple helical  
 CC collagenous domain and all 233 residues of the C-terminal non-  
 CC collagenous (NC1) domain of the alpha 3 (IV) chain. This bovine  
 CC cDNA clone was used to screen a human kidney cDNA library and a  
 CC 2.7 kb human cDNA clone (clone KMC27) was obt'd. This clone encodes  
 CC 218 residues of the NC1 domain and a portion of the 3' UTR region  
 CC of the human alpha 3(IV) chain. The COL4A3 gene localises to  
 CC chromosome 2 and therefore mutations in COL4A3 cannot be  
 CC responsible for Alport syndrome which is X-linked. An isolated  
 CC and substantially pure nt. having the sequence in AAQ96291  
 CC is claimed.

Sequence 218 AA;

Query Match 95.8%; Score 338; DB 16; Length 218;  
 Best Local Similarity 96.9%; Pred. NO. 8.8e-37;  
 Matches 62; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EFRRSPFLECHGRGTCNTYNSYSFWMASLNPERMFRKPIPTVKAGLEKITSRCQVCM 60  
 DB 155 efrrspflechgrgtrcnynsysfwlaslnpermrfrkpiptvkaglekitsrcqvc 214  
 OY 61 KKRH 64  
 ||||  
 Db 215 kkrh 218

RESULT 7  
 AAR79163  
 ID AAR79163 standard; Protein: 471 AA.  
 AC AAR79163;  
 DT 22-DEC-1995 (first entry)  
 DE Partial sequence of bovine alpha 3 chain of type IV collagen.  
 XX  
 KW Type IV collagen; alpha 3 chain.  
 XX  
 OS Bos laurus.  
 PN US5424408-A.  
 XX 13-JUN-1995.  
 PD  
 XX 30-NOV-1990; 90US-0621091.  
 PF  
 XX 30-NOV-1990; 90US-0621091.  
 PR  
 XX (UNIV ) UNIV KANSAS MEDICAL CENT.  
 PA (UYA ) UNIV YALE.  
 XX  
 PI Hudson BG, Morrison KE, Reeders ST;  
 XX WPI: 1995-262631/34.  
 DR N-PDB; AAQ96290.  
 XX

PT CDNAs encoding human or bovine alpha-3 type 4 collagen peptide(s) -  
 PT useful for detection and therapeutic removal of antibodies associated  
 PT with Goodpasture syndrome  
 PS Disclosure; Columns 5-8; 33pp; English.  
 XX  
 CC Using the PCR with primers derived from each end of the known 27 AA  
 CC residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp  
 CC bovine genomic fragment was amplified. This fragment was then used  
 CC to a bovine lens cDNA library and a 1.5 kb partial cDNA clone was  
 CC obt'd. (clone KMC15). This encodes 238 residues of the triple helical  
 CC collagenous domain and all 233 residues of the C-terminal non-



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 06:55:22 ; Search time 26.9 Seconds

(without alignments)  
53.539 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_181\_244

Perfect score: 353  
Sequence: 1 EFRA5PFLCHGRTGTCNYNS.....KAGLEKITSRCQVCMKRRH 64

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 21252 seqs, 22503292 residues

7 number of hits satisfying chosen parameters: 6

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%  
Maximum Match 100%  
Listing first 1000 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	353	100.0	218	2	US-08-399-889-25
2	353	100.0	218	3	US-09-167-364-25
3	353	100.0	218	4	US-09-439-897-4
4	323	91.5	471	2	US-08-399-889-24
5	323	91.5	471	3	US-09-167-364-24
6	323	91.5	471	4	US-09-439-897-2

#### ALIGNMENTS

RESULT 1  
US-08-399-889-25  
Sequence 25, Application US/08399889B  
Patent No. 5973120  
GENERAL INFORMATION:  
APPLICANT: Reeders, Stephen T  
APPLICANT: Morrison, Karen E  
APPLICANT: Hudson, Billy G  
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
FILE REFERENCE: 951263A  
CURRENT APPLICATION NUMBER: US/08/399, 889B  
CURRENT FILING DATE: 1995-03-07  
EARLIER APPLICATION NUMBER: 07/621091  
EARLIER FILING DATE: 1990-11-30  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 25  
LENGTH: 218  
TYPE: PRT  
ORGANISM: Human  
US-08-399-889-25

Query Match 100.0%; Score 353; DB 2; Length 218;  
Best Local Similarity 100.0%; Pred. No. 1.6e-39;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EFRA5PFLCHGRTGTCNYNSYNSFWLASLNPFRFRKPISTYKAGLEKITSRCQVCM 60  
|||||  
Db 155 EFRA5PFLCHGRTGTCNYNSYNSFWLASLNPFRFRKPISTYKAGLEKITSRCQVCM 214

QY 61 KRRH 64  
|||||  
Db 215 KRRH 218

RESULT 2  
US-09-167-364-25  
Sequence 25, Application US/09167364  
Patent No. 6007980  
GENERAL INFORMATION:  
APPLICANT: Reeders, Stephen T  
APPLICANT: Morrison, Karen E  
APPLICANT: Hudson, Billy G  
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
FILE REFERENCE: 951263B  
CURRENT APPLICATION NUMBER: US/09/167,364  
CURRENT FILING DATE: 1998-10-07  
EARLIER APPLICATION NUMBER: 08/399889  
EARLIER FILING DATE: 1995-03-07  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 25  
LENGTH: 218  
TYPE: PRT  
ORGANISM: Human  
US-09-167-364-25

Query Match 100.0%; Score 353; DB 3; Length 218;  
Best Local Similarity 100.0%; Pred. No. 1.6e-39;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EFRA5PFLCHGRTGTCNYNSYNSFWLASLNPFRFRKPISTYKAGLEKITSRCQVCM 60  
|||||  
Db 155 EFRA5PFLCHGRTGTCNYNSYNSFWLASLNPFRFRKPISTYKAGLEKITSRCQVCM 214

QY 61 KRRH 64  
|||||  
Db 215 KRRH 218

RESULT 3  
US-09-439-897-4  
Sequence 4, Application US/09439897  
Patent No. 6277558  
GENERAL INFORMATION:  
APPLICANT: Hudson, Billy G  
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides  
FILE REFERENCE: 95-1263-C  
CURRENT APPLICATION NUMBER: US/09/439,897  
CURRENT FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 65  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 218  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-439-897-4





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 06:54:09 ; Search time 31.16 Seconds

(without alignments)  
466,924 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_54\_244

Perfect score: 1052  
Sequence: 1 QRAHQDGLTGLSCLQRFPT.....KAGLEKTIISRCQYCMKRRH 191

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%  
Maximum Match 100%

Listing first 1000 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1052	100.0	1670 1	CGHU3B collagen alpha 3(IV)
2	993	94.4	26 2	CGHU3B collagen alpha 3(IV)
3	988	93.9	471 2	A39024 collagen alpha 3(IV)

## ALIGNMENTS

RESULT 1

CGHU3B collagen alpha 3(IV) chain precursor, long splice form - human  
N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form  
C:Species: Homo sapiens (man)  
C>Date: 28-Oct-1994 #sequence.revision 03-Oct-1995 #text.change 22-Jun-1999  
R:Accession: A54763; A43928; A44043; A45971; A39786  
J:Maruyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reenders, S.T.  
J. Biol. Chem. 269, 23013-23017, 1994  
A>Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression  
A:Reference number: A54763; MUID:94364994  
A:Molecule type: mRNA  
A:Residues: 1-1670 <MAP>  
A:Cross-references: GB:880031; NID:9577563; PID:9577564  
A:Experimental source: Kidney  
R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
J. Clin. Invest. 89, 592-601, 1992  
A>Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the a  
A:Reference number: A43928; MUID:92147878  
A:Accession: A43928

A:Molecule type: mRNA  
A:Residues: 1331-1524, 1', 1526-1670 <TUR>  
A:Cross-references: GB:M81379  
A:Experimental source: Kidney  
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 267, 19780-19784, 1992  
A>Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpasture antigen.  
A:Reference number: A44043; MUID:93015826  
A:Accession: A44043  
A:Molecule type: DNA  
A:Residues: 1386-1670 <OUT>  
A:Cross-references: GB:M92993; NID:9177895; PID:AAA21610.1; PID:9177896  
A>Note: sequence extracted from NCBI backbone (NCBI:P115597)  
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 269, 17358, 1994  
A:Reference number: A44738; MUID:94274734  
A:Contents: annotation; erratum; correction to intronic sequence in A44043  
R:Bernal, D.; Quinones, S.; Saus, J.  
J. Biol. Chem. 268, 12090-12094, 1993  
A>Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
A:Reference number: A45971; MUID:93280184  
A:Accession: A45971  
A>Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1427-1444 <BER>  
A>Note: sequence extracted from NCBI backbone (NCBI:P133363); sequence incorrectly id  
R:Morrison, K.E.; Maruyama, M.; Yang-Feng, T.L.; Reenders, S.T.  
Am. J. Hum. Genet. 49, 545-554, 1991  
A>Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain  
A:Reference number: A39786; MUID:91333570  
A:Accession: A39786  
A:Molecule type: mRNA  
A:Residues: 1453-1593, 'A', 1595-1670 <MOR>  
A:Cross-references: GB:S55790; NID:9234418; PID:AA819637.1; PID:9234419  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epit  
C:Genetics:  
A:Gene: GDB:COL4A3  
A:Cross-references: GDB:128351; OMIM:120070  
A:Map position: 2q36-2q37  
A:Intons: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete  
A>Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands w  
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha  
monomer trimers amino-terminal domains (with disulfide and desmosine cross-links), dimer  
er associations in the interrupted helical domain (with disulfide and desmosine cross  
C:Function:  
A:Description: minor structural component of extracellular basement membrane in kidne  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extra  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MA  
F:29-42/Domain: amino-terminal nonhelical, NHI <NHI>  
F:43-138/Region: Interrupted helical  
F:91-793/Region: cell attachment (R-G-D) motif  
F:996-998/Region: cell attachment (R-G-D) motif  
F:1154-1156/Region: cell attachment (R-G-D) motif  
F:1306-1308/Region: cell attachment (R-G-D) motif  
F:1345-1347/Region: cell attachment (R-G-D) motif  
F:1432-1434/Region: cell attachment (R-G-D) motif  
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTR1>  
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR2>  
F:31-33, 39, 41, 125, 422, 476, 479, 682, 722, 809, 1387/Disulfide bonds: Interchain #status pr  
F:253/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:1460-1548, 1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
F:1505-1511, 1616-1622/Disulfide bonds: #status predicted  
F:1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 100.0%; Score 1052; DB 1; Length 1670;  
Best Local Similarity 100.0%; Pred. No. 7.3e-93;

Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGODLTGSLCLOPRTTTPFLPCNVNDVCFASRNDYSYWLSTPALMNNAPITGR 60  
 Db 1480 ORAHGODLTGSLCLOPRTTTPFLPCNVNDVCFASRNDYSYWLSTPALMNNAPITGR 1539  
 QY 61 ALEPYISRCTVCEGPALIAVAHSQTTDIPPCPHGMISLMKGFSTIMFTSAGSEGTGALA 120  
 Db 1540 ALEPYISRCTVCEGPALIAVAHSQTTDIPPCPHGMISLMKGFSTIMFTSAGSEGTGALA 1599  
 QY 121 SPGSCLEFRASPFLCHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAAGLEKII 180  
 Db 1600 SPGSCLEFRASPFLCHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAAGLEKII 1659  
 QY 181 SRCQVCMKKRH 191  
 Db 1660 SRCQVCMKKRH 1670

LT 2

148302

collagen alpha 3(IV) chain - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 16-Feb-1997  
 C:Accession: 148302; S47278  
 R:Miner, J.H.; Sanees, J.R.  
 J. Cell Biol. 127, 879-891, 1994  
 A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequence  
 A:Reference number: A54979; MUID:95050957  
 A:Accession: 148302  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-246 <RES>  
 A:Cross-references: EMBL:235166; NID:9535197; PID:9535198  
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 94.4%; Score 993; DB 2; Length 246;

Best Local Similarity 92.7%; Pred. No. 4.6e-88;

Matches 177; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

QY 1 ORAHGODLTGSLCLOPRTTTPFLPCNVNDVCFASRNDYSYWLSTPALMNNAPITGR 60  
 Db 56 KRAHGODLTGSLCLOPRTTTPFLPCNVNDVCFASRNDYSYWLSTPALMNNAPITGR 115  
 QY 61 ALEPYISRCTVCEGPALIAVAHSQTTDIPPCPHGMISLMKGFSTIMFTSAGSEGTGALA 120  
 Db 116 ALEPYISRCTVCEGPALIAVAHSQTTDIPPCPHGMISLMKGFSTIMFTSAGSEGTGALA 175  
 QY 121 SPGSCLEFRASPFLCHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAAGLEKII 180  
 Db 176 SPGSCLEFRASPFLCHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAAGLEKII 235  
 QY 181 SRCQVCMKKRH 191  
 Db 236 SRCQVCMKKRH 246

RESULT 3

A39024

collagen alpha 3(IV) chain - bovine (fragment)

C:Species: Bos primigenius taurus (cattle)

C:Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Aug-1999

C:Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815

R:Morison, K.E.; Germino, G.G.; Reeders, S.T.

J. Biol. Chem. 266, 34-39, 1991

A:Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the

A:Reference number: A39024; MUID:91093146

A:Accession: A39024

A:Molecule type: mRNA

A:Residues: 1-471 <MOR>

A:Cross-references: EMBL:M63139; NID:9162886; PID:AAA62708.1; PID:9162887

R:Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.

J. Biol. Chem. 262, 7874-7877, 1987

A:Title: Localization of the Goodpasture epitope to a novel chain of basement membran

A:Reference number: S18432; MUID:87222419

A:Accession: S20672

A:Molecule type: protein

A:Residues: 227-228, 'X', 230-244 <BUR>

R:Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.

J. Biol. Chem. 263, 13374-13380, 1988

A:Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collag

A:Reference number: S17802; MUID:88330844

A:Accession: S17802

A:Molecule type: protein

A:Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>

R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.

J. Biol. Chem. 265, 5466-5469, 1990

A:Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of t

A:Reference number: A35167; MUID:90202779

A:Accession: A35167

A:Molecule type: protein

A:Residues: 236-258 <GUN>

R:Gunwar, S.; Ballaster, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.;

J. Biol. Chem. 266, 15318-15324, 1991

A:Title: Glomerular basement membrane. Identification of dimeric subunits of the nonc

A:Reference number: A39419; MUID:91332055

A:Accession: C39419

A:Molecule type: protein

A:Residues: 236-255 <GUN>

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication

F:1-238/Domain: collagenous (fragment) #status predicted <COL>

F:239-471/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>

F:239-353/Domain: repeat NC1 #status predicted <NC12>

F:354-471/Domain: repeat NC1 #status predicted <NC12>

F:222,238/Modified site: hydroxyproline (Pro) #status experimental

F:306-312,417-423/Disulfide bonds: #status predicted

Query Match 93.9%; Score 988; DB 2; Length 471;

Best Local Similarity 93.2%; Pred. No. 2.7e-87;

Matches 177; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 1 ORAHGODLTGSLCLOPRTTTPFLPCNVNDVCFASRNDYSYWLSTPALMNNAPITGR 60  
 Db 281 ORAHGODLTGSLCLOPRTTTPFLPCNVNDVCFASRNDYSYWLSTPALMNNAPITGR 340  
 QY 61 ALEPYISRCTVCEGPALIAVAHSQTTDIPPCPHGMISLMKGFSTIMFTSAGSEGTGALA 120  
 Db 341 ALEPYISRCTVCEGPALIAVAHSQTTDIPPCPHGMISLMKGFSTIMFTSAGSEGTGALA 400  
 QY 121 SPGSCLEFRASPFLCHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAAGLEKII 180  
 Db 401 SPGSCLEFRASPFLCHGRGTCNYNSYSFWLASLNPERMFRKPIPTVKAAGLEKII 460  
 QY 181 SRCQVCMKKR 190  
 Db 461 SRCQVCMKKR 470

Search completed: March 6, 2002, 06:54:49

Job time: 40 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2002, 06:55:24 ; Search time 18.75 Seconds  
(Without alignments)  
373.492 Million cell updates/sec

Title: US-09-543-371-10\_COPY\_54\_244  
Perfect score: 1052  
Sequence: 1 QRANGODLGTGLSCLOPFTT.....KAGELEKITSRCQYCMKKRH 191

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 90%  
Maximum Match 100%  
Listing first 1000 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1052	100.0	1670	1 CA34_HUMAN	001955 homo sapien
2	988	93.9	471	1 CA34_BOVIN	Q28084 bos taurus

## ALIGNMENTS

RESULT 1  
C7 HUMAN STANDARD; PRT; 1670 AA.  
I CA34\_HUMAN  
AC 001955;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE COLLAGEN ALPHA 3(IV) CHAIN PRECURSOR (GOODPASTURE ANTIGEN).  
GN COL4A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=94364994; PubMed=8083201;  
RA Matiyama M., Leinonen A., Mochizuki T., Tryggvason K., Reeders S.T.;  
RT "Complete primary structure of the human alpha 3(IV) collagen chain.  
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  
RT human tissues.";  
RL J. Biol. Chem. 269:23013-23017(1994).  
RN [2]  
RP REVISIONS.  
RA Leinonen A.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RN [3]  
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=93015826; PubMed=1400291;  
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;  
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the  
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
RT antigenic region at the triple helix/NC1 domain junction.";  
RL J. Biol. Chem. 267:19780-19784(1992).  
RN [4]  
RP SEQUENCE OF 1453-1670 FROM N.A.  
RX MEDLINE=9135370; PubMed=1802840;  
RA Morrison K.E., Matiyama M., Yang-Peng T.L., Reeders S.T.;  
RT "Sequence and localization of a partial cDNA encoding the human alpha  
RT 3 chain of type IV collagen.";  
RL Am. J. Hum. Genet. 49:545-554(1991).  
RN [5]  
RP SEQUENCE OF 1331-1670 FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=92147878; PubMed=1737849;  
RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
RA Pusey C.D.;  
RT "Molecular cloning of the human Goodpasture antigen demonstrates it.  
RT to be the alpha 3 chain of type IV collagen.";  
RL J. Clin. Invest. 89:592-601(1992).  
RN [6]  
RP SEQUENCE OF 1644-1670 FROM N.A.  
RC TISSUE=Kidney;  
RX Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.  
RC TISSUE=Kidney;  
RX MEDLINE=94124597; PubMed=8294492;  
RA Feng L., Xia Y., Wilson C.B.;  
RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)  
RT collagen gene. Differential expression of mRNA transcripts that  
RT predict three protein variants with distinct carboxyl regions.";  
RL J. Biol. Chem. 269:2342-2348(1994).  
RN [8]  
RP SEQUENCE OF 1-29 FROM N.A.  
RX MEDLINE=98196854; PubMed=9537506;  
RA Komota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioaka H.,  
RA Nimohiya Y.;  
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
RT 2q36.";  
RL FEBS Lett. 424:11-16(1998).  
RN [9]  
RP ALTERNATIVE SPLICING.  
RX MEDLINE=93280184; PubMed=8505332;  
RA Bernal D., Quinones S., Saus J.;  
RT "The human mRNA encoding the Goodpasture antigen is alternatively  
RT spliced.";  
RL J. Biol. Chem. 268:12090-12094(1993).  
RN [10]  
RP VARIANT PRO-1474.  
RX MEDLINE=95078827; PubMed=7987301;  
RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
RA Barrientos A., Momms L.A.H., van Oost B.A., Brunner H.G.,  
RA Reeders S.T., Smeets H.J.M.;  
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
RT recessive Alport syndrome.";  
RL Hum. Mol. Genet. 3:1269-1273(1994).  
RN [11]  
RP VARIANTS AS, AND VARIANTS.  
RX MEDLINE=21064696; PubMed=11134255;  
RA Heidet L., Arondel C., Forestier L., Cohen-Solal L., Mollet G.,  
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
RT autosomal Alport syndrome.";  
RL J. Am. Soc. Nephrol. 12:97-106(2001).  
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'

FT	SITE	791	793	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	966	998	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1154	1156	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1306	1308	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1345	1347	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1432	1434	CELL ATTACHMENT SITE (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC.. ) (POTENTIAL).
FT	MOD_RES	1435	1435	PHOSPHORYLATION (BY SIMILARITY).
FT	MOD_RES	1437	1437	PHOSPHORYLATION (BY SIMILARITY).
FT	DISULFID	1460	1551	OR 1548 (BY SIMILARITY).
FT	DISULFID	1493	1548	OR 1551 (BY SIMILARITY).
FT	DISULFID	1505	1511	BY SIMILARITY.
FT	DISULFID	1570	1665	OR 1662 (BY SIMILARITY).
FT	DISULFID	1604	1662	OR 1665 (BY SIMILARITY).
FT	DISULFID	1616	1622	BY SIMILARITY.
FT	VARSPLIC	1586	1670	FTSAGSBGTGALASPGSCLEFPFRASPFLCHGRGCNCNYS NSTSYWLASLWPEEMFKPIPLSTVYKAGELEKIISRCQCMK RKH -> KAVSINCEESWGIRKNKNSLSGVHEERTLKTKTA ELVEFLITNKXMEHVAI (IN ISOFORM 2).
FT	VARSPLIC	1488	1670	GTIGSLORFTTMTPLFCNDVCNPFASRDYSWTSLPALMNNAPITGR MPMNAPITGRALPELYISRCTYCEGPAIAIVAHSGTTDIP CPHGWSILMKGFSPIMFTSAGSBGTGALASGSCLEEFRA SPFLCHGRGCNCNYSNSYSFWLASLNPRMRKRPIPTSVK AGELEKIISRCQCMKKRH -> DALFVKVLASP (IN ISOFORM 3).
FT	VARIANT	43	43	G -> R. /FTid=VAR_011202.
FT	VARIANT	162	162	G -> E. /FTid=VAR_011203.
FT	VARIANT	297	297	G -> E (IN AS).
Query Match		100.0%	Score 1052;	DB 1;
Best Local Similarity		100.0%;	Pred. No. 3, se-96;	Length 1670;
Matches 191;	Conservative	0;	Mismatches 0;	Indels 0;
				Gaps 0.
OY	1	QRANGODLGTLTGSLORFTTMTPLFCNVNDVCNPFASRDYSWTSLPALMNNAPITGR	60	
Db	1480	QRANGODLTGTLGSLOFTTMTPLFCNVNDVCNPFASRDYSWTSLPALMNNAPITGR	1539	
OY	61	ALEPYISRCTYCEGPAIAIVAHSGTTDIPPCPHGWISLMKGFSIFTMSAGSBGTGALA	120	
Db	1540	ALEPISNCTYCEGPAIAIVAHSGTTDIPPCPHGWISLMKGFSIFTMSAGSBGTGALA	1599	
OY	121	SPGSCLEFPFRASPFLCHGRGCNCNYSNSYSFWLASLNPRMRKRPIPTSVKAGELKII	180	
Db	1600	SPGSCLEFPFRASPFLCHGRGCNCNYSNSYSFWLASLNPRMRKRPIPTSVKAGELKII	1659	
OY	181	SRQCVMCKRRH	191	
Db	1660	SRQCVMCKRRH	1670	
RESULT	2			
CA34_BOVIN		STANDARD:	PRT:	471 AA.
ID	CA34_BOVIN			
AC	028084:			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	COLLAGEN ALPHA 3(IV) CHAIN (FRAGMENT).			
GN	COLA43.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;			
OX	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lens;			
RX	MEDLINE=91093146; PubMed=1985905;			
RA	Morrison K.E., Gernino G.G., Reeders S.T.;			
RT	"Use of the polymerase chain reaction to clone and sequence a cDNA encoding the bovine alpha 3 chain of type IV collagen.";			

RL J. Biol. Chem. 266:34-39(1991).  
 RN [2] SEQUENCE OF 227-258.  
 RP TISSUE-Kidney;  
 RX MEDLINE=90202779; PubMed=2318822;  
 RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;  
 RT "Glomerular basement membrane. Identification of a fourth chain,  
 alpha 4, of type IV collagen.";  
 RL J. Biol. Chem. 265:5466-5469(1990).  
 RN [3] SEQUENCE OF 227-254.  
 RP MEDLINE=88330844; PubMed=3417661;  
 RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;  
 RT Identification of the Goodpasture antigen as the alpha 3(IV) chain  
 of collagen IV.";  
 RL J. Biol. Chem. 263:13374-13380(1988).  
 RN [4] SEQUENCE OF 227-244.  
 RP MEDLINE=87222419; PubMed=2438283;  
 RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,  
 R. Hudson B.G.;  
 RT Localization of the Goodpasture epitope to a novel chain of basement  
 membrane collagen.";  
 RL J. Biol. Chem. 262:7874-7877(1987).  
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/  
 NIDOGEN.  
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOMERISMS, ALPHA 1(IV)-  
 ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 IV COLLAGENS.  
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 DR EMBL: M63139; AAA62708.1;  
 DR InterPro: IPR001442; C4.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01413; C4; 2.  
 DR Pfam: PF01391; Collagen; 4.  
 DR PRODOM: PD003923; C4; 2.  
 DR SMART: SM00111; C4; 2.  
 KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
 FT NON\_TER 1  
 FT DOMAIN <1 238  
 FT DOMAIN 239 471  
 FT SITE 106 108  
 FT MOD\_RES 232 232  
 FT MOD\_RES 238 238  
 FT DISULFID 261 352  
 FT DISULFID 294 349  
 FT DISULFID 306 312  
 FT DISULFID 371 466  
 FT DISULFID 405 463  
 OR 463 (BY SIMILARITY).  
 OR 466 (BY SIMILARITY).

FT DISULFID 417 423 BY SIMILARITY.  
 FT CONFLICT 253 253 S -> Y (IN REF. 3).  
 SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;  
 Query Match 93.98; Score 988; DB 1; Length 471;  
 Best Local Similarity 93.28; Pred. No. 1.9e-90;  
 Matches 177; Conservative 9; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 ORAHGDLGTLGSLQRTMPFLFCNVNDVCNFAANDYSYWLSTPALPMNNAPTGR 60  
 Db :|||||  
 Db 281 EQAHGDLGTLGSLQRTMPFLFCNVNDVCNFAANDYSYWLSTPALPMNNAPTGR 340  
 QY 61 ALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWMISLMKGFSTFMTSAGSEGTGALA 120  
 Db :|||||  
 Db 341 ALPEYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWMISLMKGFSTFMTSAGSEGTGALA 400  
 QY 121 SPSCLEEFRAFPLECHGCTGCTNYNSYSFWLASLNPERRMRKPIPSYVKGAELEKII 180  
 Db :|||||  
 Db 401 SPSCLEEFRAFPLECHGCTGCTNYNSYSFWLASLNPERRMRKPIPSYVKGAELEKII 460  
 QY 181 SRCOVCMKKR 190  
 Db :|||||  
 Db 461 SRCOVCMKKR 470

Search completed: March 6, 2002, 07:05:08  
 Job time: 584 sec

THIS PAGE BLANK (USPTO)